

数学建模培训教材

Methods and Applications
of Mathematical Modeling

数学建模 方法及其应用

解放军信息工程大学 韩中庚 编著



高等教育出版社
HIGHER EDUCATION PRESS

数学建模培训教材

数学建模方法及其应用

Methods and Applications of Mathematical Modeling

解放军信息工程大学 韩中庚 编著

高等教育出版社

内容简介

本书主要是根据“数学建模”课程的教学和数学建模竞赛培训活动的实际需要,以及作者多年从事相关工作的实践经验和体会编写而成的,从内容上突出体现了“广、浅、新、用”的现代应用特点。

主要内容包括量纲分析、集合分析、微分方程、差分方程、插值与拟合、层次分析、概率分布、数理统计、回归分析、线性规划、整数规划、非线性规划、动态规划、排队论、对策论、随机性决策分析、多目标决策分析、图论、模糊数学和灰色系统分析等 20 大类数学建模方法,每一种方法都有相应的应用案例分析及参考案例。最后附有历年中国大学生数学建模竞赛和美国大学生数学建模竞赛的问题,以及 MATLAB 的使用简介。

本书可作为专科生、本科生、研究生的“数学建模”课程教材外,还特别适用于数学建模竞赛的培训教材,以及供从事应用研究的工程技术人员参考之用。

图书在版编目(CIP)数据

数学建模方法及其应用 / 韩中庚编著. —北京:高等教育出版社,2005.6

ISBN 7-04-016328-4

I. 数... II. 韩... III. 数学模型 - 高等学校 - 教材 IV. 022

中国版本图书馆 CIP 数据核字(2005)第 042978 号

策划编辑 徐可 责任编辑 张耀明 封面设计 王凌波 责任绘图 朱静
版式设计 张岚 责任校对 殷然 责任印制 杨明

出版发行 高等教育出版社
社址 北京市西城区德外大街 4 号
邮政编码 100011
总机 010-58581000
经销 北京蓝色畅想图书发行有限公司
印刷 国防工业出版社印刷厂

购书热线 010-58581118
免费咨询 800-810-0598
网址 <http://www.hep.edu.cn>
<http://www.hep.com.cn>
网上订购 <http://www.landracom.com>
<http://www.landracom.com.cn>

开本 787×960 1/16
印张 30
字数 560 000

版次 2005 年 6 月第 1 版
印次 2005 年 6 月第 1 次印刷
定价 31.10 元

本书如有缺页、倒页、脱页等质量问题,请到所购图书销售部门联系调换。

版权所有 侵权必究

物料号 16328-00

序 言

虽然,数学建模(Mathematical Modeling)作为一个词汇问世,不过是近几十年的事情,可是,在数学产生、发展的历史长河中,人们用建立数学模型的办法解决那些需要数量规律的现实问题,并获得巨大的成功,是不乏先例的。进入20世纪以来,数学的应用不仅在它的传统领域——工程技术、经济建设发挥着越来越重要的作用,而且不断地向一些新的领域渗透,形成了许多交叉学科——计量经济学、人口控制论、生物数学、地质数学等等。数学与计算机技术相结合,形成了一种普遍的、可以实现的技术——数学技术,成为当代高新技术的重要组成部分。“高技术本质上是数学技术”的观点已被越来越多的人所接受。在高科技,特别是计算机技术迅速发展的今天,计算和建模正在成为数学科学向数学技术转化的主要途径。

教育必须跟踪、反映并满足社会的实际需要,数学建模进入大学课堂,既顺应时代发展的潮流,也符合教育改革的要求。对于数学教育而言,既应该让学生掌握准确快捷的计算方法和严密的推理论证,也需要培养用数学工具表达、分析与解决实际问题的意识和能力,传统的数学教学体系和内容无疑偏重于前者,开设数学建模课程则是加强后者的一种尝试。

1992年,全国大学生数学建模竞赛一出现就受到广大同学的欢迎和教育界的重视。十几年来,数学建模课堂教学和竞赛活动相互促进、相得益彰,其内涵之丰富和规模之庞大,在数学教学乃至整个大学教育中都是少见的。许多学校还经常举办校内的建模竞赛,再加上参加培训的同学数倍于参加竞赛的同学,这项活动的受益面是相当可观的。

许多中青年教师常年辛勤工作在数学建模教学和竞赛的第一线,以精湛的指导能力,一丝不苟、认真细致的工作态度,任劳任怨、不计报酬的奉献精神,得到同学们由衷的称赞,为数学建模活动的发展做出了巨大贡献。这本书的作者就是他们中突出的一位。

我们提倡广大同学通过数学建模课的学习和参加数学建模竞赛有关的课外活动,使更多同学参与到数学建模中来,而且能使他们从中都能受益,特别是通过学习和训练能够更多地掌握数学建模的知识和用数学工具分析、解决实际问题的方法,进一步提高数学建模的意识和能力。目前,适合同学们学习和参加竞赛用的辅导材料还不够丰富。

这本书的作者在多年讲授数学建模课程及组织、辅导竞赛活动的同时,做了

II 序言

不少与建模有关的实际应用课题。他的这些经历使这本书作为数学建模教学和竞赛的辅导材料至少有两个显著特点,一是比较全面、系统和简练地介绍了数学建模常用的各种方法,二是每种方法都附有生动的实例,并且这些实例大多数来自作者自己的研究工作,这在我国出版的数学建模图书中是比较少见的。

希望这本书的出版能让更多的教师和同学在数学建模活动中得到有益的启迪和更大的收获。

姜启源

清华大学数学科学系

2004年10月

前言

目前,数学建模现已成为大学校园里的一个热门话题,参加数学建模竞赛活动在广大的学生中已成为追求的时尚,这项工作越来越受到各级教育管理部门领导的关心和重视,也得到了科技界和教育界的普遍关注。这主要是数学建模的教学和竞赛活动在综合能力、创新意识、科研素质的培养过程中的作用和地位已被越来越多的人所认识,它的实际效果正在源源不断地显现出来,“数学建模的人才”和“数学建模的能力”正在实际工作中发挥着积极的重要作用。数学建模工作的发展前景是无限的,数学建模的人才大有作为。

一年一度的全国大学生数学建模竞赛是为了推动高校的教学改革、更好地培养学生解决实际问题的能力和创造精神,是由教育部高等教育司直接组织领导,面向全国高等院校的规模最大、参与院校最多、涉及面最广的一项科技竞赛活动。自1992年举办第一届竞赛以来,参赛队数平均每年以20%左右的速度增加,2004年已到了724所院校的6881个参赛队。也正是由于数学建模活动的开展,有力地推动了大学数学改革的开展,并已取得了显著的成果,全国范围内在校研究生的数学建模竞赛也正在悄然兴起,并乘全国大学生数学建模竞赛的东风,大有飞跃发展之势,相信不久的将来这项面向全面高等院校研究生的科技竞赛活动像全国大学生数学建模竞赛一样,健康蓬勃地发展壮大。

作者在多年从事“数学建模”课程的教学和组织指导数学建模竞赛工作的基础上,博览了国内现有的大部分相关资料,并吸取了各自所长,编著了这本教材,其内容是作者多年从事相关工作的积累和总结,书中所讲授的数学建模方法是目前应用最广泛的方法,也是数学建模工作应该掌握的方法,所引用的案例大部分是作者自己的工作,当然,限于篇幅不可能把所有有关的内容都写进书里,但在书中充分体现了“广、浅、新、用”的现代应用特点。

本书可以作为专科生、本科生的“数学建模”课程教材,在规定的学时内选择相应的教学内容;作为大学生、研究生参加数学建模竞赛的培训教材都是必须要掌握的内容;也可以作为相关专业研究生和技术人员从事应用研究工作参考之用。

本书之所以能够献给读者,主要是得到了解放军信息工程大学、信息工程学院各级领导和学院机关的关心与支持,以及全国大学生数学建模竞赛组委会的资助和专家的指导帮助;特别是清华大学姜启源教授在百忙之中审阅了全部书稿,多次提出指导性修改意见,并为书作序;解放军信息工程大学杨士杰教授也

II 前言

审阅了本书的初稿,提出了很好的建议;我的几位研究生杜建平、田园、陈廷、魏福林等也为本书做了许多的工作。在此,对他们的关心、支持和帮助一并表示衷心的感谢。在这里还要感谢多年来同我一起并肩战斗过的同事们,以及所有参加过数学建模竞赛活动的同学们,正是有了他(她)们的支持和合作,才给了我学习、提高的发展空间和机遇。我是数学建模的受益者,誓将数学建模的事业发扬光大,让更多的人从中受益,为培养更多高素质、具有创新能力的应用人才做出应有的贡献。

限于作者的才疏学浅,书中肯定有许多不妥之处,敬请批评指正。

韩中庚

2005年2月于郑州

目 录

| | |
|-------------------------|----|
| 第一章 引言 | 1 |
| 1.1 数学建模的作用和地位 | 1 |
| 1.2 什么是数学模型? | 4 |
| 1.3 数学模型无处不在 | 5 |
| 1.4 数学建模的方法和步骤 | 15 |
| 1.5 数学建模与能力的培养 | 18 |
| 1.6 参考案例与参考文献 | 19 |
| 第二章 两种初等分析方法 | 21 |
| 2.1 量纲分析方法 | 21 |
| 2.2 空间点热源的扩散问题 | 23 |
| 2.3 集合分析方法 | 25 |
| 2.4 合理分派与会成员问题 | 27 |
| 2.5 参考案例与参考文献 | 36 |
| 第三章 微分方程方法 | 37 |
| 3.1 微分方程的一般理论 | 37 |
| 3.2 微分方程的平衡点及稳定性 | 40 |
| 3.3 战争的预测与评估问题 | 42 |
| 3.4 SARS 传播问题 | 45 |
| 3.5 参考案例与参考文献 | 54 |
| 第四章 差分方程方法 | 55 |
| 4.1 常系数线性差分方程 | 55 |
| 4.2 差分方程的平衡点及其稳定性 | 57 |
| 4.3 连续模型的差分方法 | 58 |
| 4.4 最优捕鱼问题 | 64 |
| 4.5 参考案例与参考文献 | 68 |
| 第五章 插值与拟合方法 | 69 |
| 5.1 一般插值方法 | 69 |
| 5.2 样条函数插值方法 | 72 |

| | | |
|------------|---------------------|------------|
| 5.3 | B 样条函数插值方法 | 76 |
| 5.4 | 最小二乘拟合方法 | 80 |
| 5.5 | 黄河小浪底调水调沙问题 | 82 |
| 5.6 | 参考案例与参考文献 | 90 |
| 第六章 | 层次分析方法 | 91 |
| 6.1 | 层次分析的一般方法 | 91 |
| 6.2 | 一类择优排序问题 | 96 |
| 6.3 | 合理分配住房问题 | 100 |
| 6.4 | 参考案例与参考文献 | 107 |
| 第七章 | 概率分布方法 | 108 |
| 7.1 | 排列与组合 | 108 |
| 7.2 | 事件与概率 | 109 |
| 7.3 | 随机变量与分布函数 | 111 |
| 7.4 | 常用的概率分布及数字特征 | 114 |
| 7.5 | 足球门的危险区域问题 | 116 |
| 7.6 | 参考案例与参考文献 | 123 |
| 第八章 | 数理统计方法 | 125 |
| 8.1 | 统计方法的基本概念 | 125 |
| 8.2 | 参数估计法 | 127 |
| 8.3 | 方差分析法 | 130 |
| 8.4 | 相关分析方法 | 133 |
| 8.5 | 最优评卷问题 | 134 |
| 8.6 | 参考案例与参考文献 | 141 |
| 第九章 | 回归分析方法 | 143 |
| 9.1 | 一元线性回归方法 | 143 |
| 9.2 | 多元线性回归方法 | 148 |
| 9.3 | 回归模型的选择方法 | 153 |
| 9.4 | 回归模型的正交化设计方法 | 155 |
| 9.5 | 多重共线性与有偏估计方法 | 157 |
| 9.6 | 沼气的生成问题 | 159 |
| 9.7 | 参考案例与参考文献 | 162 |
| 第十章 | 线性规划方法 | 164 |
| 10.1 | 线性规划的模型 | 164 |

| | | |
|---------------------------|-------------------------|------------|
| 10.2 | 线性规划解的概念与理论 | 166 |
| 10.3 | 线性规划的求解方法 | 167 |
| 10.4 | 线性规划的对偶问题 | 169 |
| 10.5 | 线性规划的灵敏度分析 | 171 |
| 10.6 | 南水北调水指标的分配问题 | 173 |
| 10.7 | 参考案例与参考文献 | 178 |
| 第十一章 整数规划方法 | | 180 |
| 11.1 | 整数规划的模型 | 180 |
| 11.2 | 整数规划的分枝定界法 | 181 |
| 11.3 | 整数规划的割平面法 | 183 |
| 11.4 | 0-1 整数规划 | 185 |
| 11.5 | 指派问题的匈牙利方法 | 188 |
| 11.6 | 招聘公务员问题 | 190 |
| 11.7 | 参考案例与参考文献 | 201 |
| 第十二章 非线性规划方法 | | 203 |
| 12.1 | 非线性规划的基本概念 | 203 |
| 12.2 | 无约束非线性规划的解法 | 205 |
| 12.3 | 带有约束条件的非线性规划 | 208 |
| 12.4 | 非线性规划的制约函数法 | 211 |
| 12.5 | 奶制品的加工计划问题 | 214 |
| 12.6 | 参考案例与参考文献 | 218 |
| 第十三章 动态规划方法 | | 220 |
| 13.1 | 动态规划的基本概念和基本方程 | 220 |
| 13.2 | 动态规划的求解方法 | 224 |
| 13.3 | 动态规划方法的应用 | 225 |
| 13.4 | 选拔队员与组队问题 | 229 |
| 13.5 | 参考案例与参考文献 | 236 |
| 第十四章 排队论方法 | | 238 |
| 14.1 | 排队论的基本概念 | 238 |
| 14.2 | 到达时间的间隔分布和服务时间的分布 | 240 |
| 14.3 | 单服务台的排队模型 | 243 |
| 14.4 | 多服务台的排队模型 | 248 |
| 14.5 | 排队系统的最优化问题 | 251 |

| | | |
|-------------|------------------------|------------|
| 14.6 | 校园网的设计和调节收费问题 | 253 |
| 14.7 | 参考案例与参考文献 | 256 |
| 第十五章 | 对策论方法 | 258 |
| 15.1 | 对策论的基本概念 | 258 |
| 15.2 | 矩阵对策的概念和理论 | 259 |
| 15.3 | 矩阵对策的解法 | 264 |
| 15.4 | 双矩阵对策 | 265 |
| 15.5 | 玫瑰有约问题 | 269 |
| 15.6 | 参考案例与参考文献 | 277 |
| 第十六章 | 随机性决策分析方法 | 279 |
| 16.1 | 随机性决策问题的基本概念 | 279 |
| 16.2 | 效用函数理论 | 281 |
| 16.3 | 彩票中的数学问题 | 285 |
| 16.4 | 参考案例与参考文献 | 296 |
| 第十七章 | 多目标决策分析方法 | 298 |
| 17.1 | 多目标决策分析的基本概念 | 298 |
| 17.2 | 多目标决策问题的非劣解 | 301 |
| 17.3 | 多目标群决策问题的解 | 303 |
| 17.4 | 股份制公司的综合投资问题 | 306 |
| 17.5 | 参考案例与参考文献 | 310 |
| 第十八章 | 图论方法 | 312 |
| 18.1 | 图 | 312 |
| 18.2 | 树 | 314 |
| 18.3 | 遍历 | 315 |
| 18.4 | 匹配 | 316 |
| 18.5 | 图矩阵 | 317 |
| 18.6 | 婚配问题 | 318 |
| 18.7 | 锁具装箱问题 | 320 |
| 18.8 | 参考案例与参考文献 | 321 |
| 第十九章 | 模糊数学方法 | 322 |
| 19.1 | 模糊数学的基本概念 | 322 |
| 19.2 | 模糊关系与模糊矩阵 | 326 |
| 19.3 | 模糊聚类分析方法 | 328 |

| | | |
|-------------|--------------------------------------|------------|
| 19.4 | 模糊模式识别方法 | 333 |
| 19.5 | 模糊综合评判方法 | 335 |
| 19.6 | 中介服务机构的信誉评估问题 | 338 |
| 19.7 | 参考案例与参考文献 | 343 |
| 第二十章 | 灰色系统分析方法 | 345 |
| 20.1 | 灰色系统分析的基本概念 | 345 |
| 20.2 | 灰色模型 GM | 350 |
| 20.3 | 灰色预测 | 353 |
| 2*0.4 | 灰色决策 | 355 |
| 20.5 | SARS 疫情对某些经济指标影响问题 | 357 |
| 20.6 | 参考案例与参考文献 | 362 |
| 附录 A | 中国大学生数学建模竞赛题(1992—2004) | 364 |
| 附录 B | 美国大学生数学建模竞赛题(1985—2005) | 411 |
| 附录 C | MATLAB 的使用简介..... | 447 |

第一章 引言

众所周知,21 世纪是知识经济的时代,所谓知识经济是以现代科学技术为核心,建立在知识和信息的生产、存储、使用和消费之上的经济;是以智力资源为第一生产力要素的经济;是以高科技产业为支柱产业的经济.知识创新和技术创新是知识经济的基本要求和内在动力,培养高素质、复合型的创新人才是时代发展的需要.创新人才主要是指具有较强的创新精神、创新意识和创新能力,并能够将创造能力转化为创造性成果的高素质人才.培养创新人才,大学教育是关键,而大学的数学教育在整个大学教育,乃至在人才的培养中都起着重要的奠基作用.正如著名的数学家王梓坤院士所说:“今天的数学兼有科学和技术两种品质,数学科学是授人以能力的技术.”数学作为一门技术,现已经成为一门能够普遍实施的技术,也是未来所需要的高素质创新人才必须要具备的一门技术.随着知识经济发展的需要,创新人才的供需矛盾日趋突现,这也是全社会急呼教学改革的根本所在.因此,现代大学数学教育的思想核心就是在保证打牢学生基础的同时,力求培养学生的创新意识与创新能力、应用意识与应用能力.也就是大学数学教育应是基于传授知识、培养能力、提高素质于一体的教育理念之下的教学体系.数学建模活动是实现这一改革目标的有效途径,也正是数学建模活动为大学的数学教学改革打开了一个突破口,近几年的实践证明,这一改革方向是正确的,成效是显著的.

1.1 数学建模的作用和地位

我们培养人才的目的是为了服务于社会、应用于社会,促进社会的进步和发展.而社会实际中的问题是复杂多变的,量与量之间的关系并不明显,并不是套用某个数学公式或只用某个学科、某个领域的知识就可以圆满解决的,这就要求我们培养的人才应有较高的数学素质.即能够从众多的事物和现象中找出共同的、本质的东西,善于抓住问题的主要矛盾,从大量数据和定量分析中寻找并发现规律,用数学的理论和数学的思维方法以及相关知识去解决,从而为社会服务.基于此,我们认为定量分析和数学建模等数学素质是知识经济时代人才素质的一个重要方面,是培养创新能力的一个重要方法和途径.因此,开展数学建

模活动将会在人才培养的过程中有着重要的地位和起着重要的作用。

1.1.1 数学建模的创新作用

数学科学在实际中的重要地位和作用已普遍地被人们所认识,它的生命力正在不断地增强,这主要是来源于它的应用地位. 各行各业和各科学领域都在运用数学,或是建立在数学基础之上的,正像人们所说的“数学无处不在”已成为不可争辩的事实. 特别是在生产实践中运用数学的过程就是一个创造性的过程,成功应用的核心就是创新. 我们这里所说的创新是指科技创新,所谓的科技创新主要是指在科学技术领域的新发明、新创造. 即发明新事物、新思想、新知识和新规律;创造新理论、新方法和新成果;开拓新的应用领域、解决新的问题. 大学是人才培养的基地,而创新人才培养的核心是创新思想、创新意识和创新能力的培养. 传统的教学内容和教学方法显然不足以胜任这一重任,数学建模本身就是一个创造性的思维过程,从数学建模的教学内容、教学方法,以及数学建模竞赛活动的培训等都是围绕着一个培养创新人才的核心这个主题内容进行的,其内容取材于实际、方法结合于实际、结果应用于实际. 总之,知识创新、方法创新、结果创新、应用创新无不在数学建模的过程中得到体现,这也正是数学建模的创新作用所在.

1.1.2 数学建模的综合作用

对于我们每一个教数学基础课的教师来说,在上第一堂课的时候,按惯例都会讲一下课程的重要性,一方面要强调课程的基础性作用;另一方面,免不了都要说它在实际中有多么重要的应用价值等等. 对大多数学生来说,可能对这门课程在实际中的应用更感兴趣,但是,往往等到课程上完了以后,经常是让这些学生大失所望,主要是因为他们没有看到课程在实际中的应用,仅仅是做了几道简单的应用题而已. 学生免不了就会质问教师:“你既然说本课程在实际中有重要的应用,那么为什么不教我们如何应用本课程的知识来解决实际问题呢?”这个问题对一般的基础课教师可能是难以明确回答的,原因是单学科的知识能够解决的实际问题是很少的,尤其是对于某些基础数学课程而言更是如此. 而学习了数学建模以后,这个问题就不存在了,因为数学建模就是综合运用所掌握的知识和方法,创造性地分析解决来自于实际中的问题,而且不受任何学科和领域的限制,所建立的数学模型可以直接应用于实际中去,这是数学建模的综合作用之一.

另一方面,数学建模的工作是综合性的,所需要的知识和方法是综合性的,所研究的问题是综合性的,所需要的能力当然也是综合性的. 数学建模的教学就是向学生传授综合的数学知识和方法,培养综合运用所掌握的知识和方法来分

析问题、解决问题的能力.结合数学建模的培训和参加建模竞赛等活动,来培养学生丰富灵活的想像能力、抽象思维的简化能力、一眼看穿的洞察能力、与时俱进的开拓能力、学以致用应用能力、会抓重点的判断能力、高度灵活的综合能力、使用计算机的动手能力、信息资料的查阅能力、科技论文的写作能力、团结协作的攻关能力等等.数学建模就是将这些能力有机地结合在一起,形成了一种超强的综合能力,我们可称之为“数学建模的能力”.这就是21世纪所需要的高素质人才应该具备的能力,我们可以断言,谁具备了这种能力,必将会大有作为.

1.1.3 数学建模的桥梁地位

传统的教学内容和方法的一个最主要的问题就是理论联系实际不够密切,甚至相脱节,以至于在社会上出现了学数学没有用的一种观点,并且产生了一定社会效应.一段时间内,一些学校的数学课时被压缩,一般院校的数学系的生源质量在下降,甚至短缺,使得一些数学系的生存能力发生危机.从而,导致了一些院校的数学系不得不改变自己的培养方向和专业设置,有的合并、有的改名,一时间如雨后春笋般地诞生了许多“信息科学与计算”、“信息与计算”、“计算机与数学”等等时髦的专业.或许这也是时代发展、与时俱进的结果吧!我们认为,关键的问题还是数学有用与数学无用的对立矛盾.在中国改革开放以后,国民经济飞速发展的时期,如果数学不能为此做出贡献,那么,被人误认为数学无用应属自然.为此,数学教学改革的呼声强烈,也势在必行.现在教学改革的春风吹遍中华大地,数学教学改革的硕果累累,但成功之作无不与数学建模有关,也正是数学建模为中国数学的发展带来了生机和希望,通过“数学建模”这座无形的桥梁使得数学在工程上、生活中都得到实际的应用,这是数学建模的桥梁作用之一.

另一方面,现有的科技人才可以分为工程应用与理论研究两大类,从某种意义上讲,工程与理论存在着客观的对立.特别是工程与数学、工程师与数学家之间在处理问题的方式方法上都客观地存在一些不同或对立的观点,于是两者之间在具体问题上缺乏共同的沟通语言.对于数学建模和数学建模的人才可以在工程与数学、工程师与数学家之间架起一座桥梁,能在两者之间建立起共同语言,使沟通无限.因为数学建模的人才具有一种特有的能力——“双向翻译能力”,即可以将实际问题简化抽象为数学问题——建立数学模型;利用计算机等工具求解数学模型,再将求解结果返回到实际中去,并用来分析解释实际问题.这就使得工程与数学有机地结合在一起,工程师与数学家之间可以无障碍地沟通与合作,这也是使得近些年来能起这种桥梁作用的数学建模和数学建模人才备受欢迎的主要原因.

钱学森说:“信息时代高技术的竞争本质上是数学技术的竞争.”换言之,高

技术的发展的关键是数学技术的发展,而数学技术与高技术结合的关键就是数学模型.数学模型就像一把金钥匙打开了高技术的道道难关,任何一项技术的发展都离不开数学模型,甚至技术水平的高低取决于数学模型的优劣.

1.2 什么是数学模型?

1.2.1 原型与模型

原型与模型是一对对偶体,原型是指人们在现实世界里关心、研究或者从事生产、管理的实际对象.而模型是指为了某个特定目的将原型的某一部分信息简缩、提炼而构造的原型替代物.模型不是原型,既简单于原型,又高于原型.例如,大家熟知的飞机模型,虽然在外观上比飞机原型简单,而且也不一定会飞,但是它很逼真,也足以让人想像飞机在飞行的过程中机翼的位置与形状的影响和作用.一个城市的交通图是该城市(原型)的模型,看模型比看原型清楚得多,此时城市的人口、道路、车辆、建筑物的形状等都不重要.但是,城市的街道、交通线路和各单位的位置等信息都一目了然,这比看原型清楚得多.模型可以分为形象模型和抽象模型,抽象模型最主要的就是数学模型.

1.2.2 数学模型

当一个数学结构作为某种形式语言(即包括常用符号、函数符号、谓词符号等符号集合)解释时,这个数学结构就称为数学模型.换言之,数学模型可以描述为:对于现实世界的一个特定对象,为了一个特定目的,根据特有的内在规律,做出一些必要的简化假设,运用适当的数学工具得到的一个数学结构.也就是说,数学模型是通过抽象、简化的过程,使用数学语言对实际现象的一个近似的刻画,以便于人们更深刻地认识所研究的对象.

数学模型并不是新的事物,自从有了数学,也就有了数学模型.即要用数学去解决实际问题,就一定要使用数学的语言、方法去近似地刻画这个实际问题,这就是数学模型.事实上,人所共知的欧几里得几何、微积分、柯西积分公式、万有引力定律、能量转换定律、广义相对论等都是非常好的数学模型.我们设想,如果现在没有这些数学模型,那么,世界将是什么样子.

实际中,能够直接使用数学方法解决的实际问题是不多的,然而,应用数学知识解决实际问题的第一步就是通过实际问题本身,从形式上杂乱无章的现象中抽象出恰当的数学关系,也就是构建这个实际问题的数学模型,其过程就是数学建模的过程.

1.2.3 数学模型与数学

数学模型与数学是不完全相同的,主要体现在三个方面:

(1) 研究内容: 数学主要是研究对象的共性和一般规律,而数学模型主要是研究对象的个性(针对性)和特殊规律.

(2) 研究方法: 数学的主要研究方法是演绎推理,即按照一般原理考察特定的对象,导出结论.而数学模型的主要研究方法是归纳加演绎,归纳是依据个别现象推断一般规律.归纳是演绎的基础,演绎是归纳的指导.即数学模型是将现实对象的信息加以翻译、归纳的结果,经过求解、演绎,得到数学上的解答,再经过翻译回到现实对象,给出分析、预报、决策、控制的结果.

(3) 研究结果: 数学的研究结果被证明了就一定是正确的,而数学模型的研究结果被证明了未必一定正确,这是因为与模型的简化和模型的假设有关,因此,对数学模型的研究结果必须接受实际的检验.

然而,鉴于数学模型与数学的关系和区别,我们评价一个数学模型优劣的标准主要是:模型是否有一定的实际背景、假设是否合理、推理是否正确、方法是否简单、论述是否深刻等等.

1.3 数学模型无处不在

目前,数学的应用已渗透到了各个领域,或者说各行各业日益依赖于数学,在人们的日常生活的各种活动中,数学无处不在.也就是说在数学发展的进程中,无时无刻无不下数学模型的烙印,在数学应用的各个领域无处没有数学模型的身影.基于数学模型的广泛应用,我们现在可以说:“数学模型在无处不在了,”人人都会接触到它.例如:生活中的合理投资问题、养老保险问题、住房公积金问题、新技术的传播问题、流言蜚语的传播问题、传染病的流行问题、语言学中用词变化问题、人口的增长问题、减肥与增肥问题以及各种资源的管理问题等等,下面给出几个简单的例子.

1.3.1 流言蜚语(或小道消息)传播问题

假设某地区的总人口为 N ,在短期内不变, $x(t)$ 表示知道消息的人数所占的百分比,初始时刻的百分比为 $x_0 < 1$,传播率为 h ,则可以建立数学模型为

$$\begin{cases} \frac{dx}{dt}N = hN(1-x), \\ x(0) = x_0, \end{cases}$$

求解易知 $x(t) = (x_0 - 1)e^{-ht} + 1$,且 $\lim_{t \rightarrow +\infty} x(t) = 1$,显然是不符合实际情况的,实

际情况是未知者会从传播中得知,传播率为 h ,而有一部分人虽知消息,但不轻信,不去传播,于是可设不传播率为 r ,则数学模型为

$$\begin{cases} \frac{dx}{dt}N = N[h - (h+r)x], \\ x(0) = x_0, \end{cases}$$

求解得

$$x(t) = \left(x_0 - \frac{h}{h+r}\right)e^{-(h+r)t} + \frac{h}{h+r},$$

于是有

$$\lim_{t \rightarrow +\infty} x(t) = \frac{h}{h+r} < 1.$$

此表明随时间的增长,消息慢慢地会淡化,逐步被人遗忘,是符合实际情况的.

1.3.2 借贷买房(或购物)问题^[5]

曾有一家报纸刊登一则广告称:“对于大多数工薪阶层的人士来说,想买房,简直是天方夜谭.现在有这样一栋住宅楼,每套只需自备款七万元,其余由公司贷款,可分期付款,每月只需付 800 元,十年还清.那么,这对您还有什么问题呢!”

现在的问题是:这房子究竟值多少钱,即如果一次付款要付多少钱?如果没有能力一次付款,实际上,相当于借了多少钱?为什么要每月付 800 元?

试根据广告所提供的信息和银行的贷款利率,对上述问题进行研究,供购房者参考.

1. 一般问题

假设房子的总价为 M 元,买者需借 A_0 元,月利率为 R ,借期为 N 个月,每月付 x 元,到第 n 个月欠款 A_n 元,则第 $n+1$ 个月(含利息)欠款

$$A_{n+1} = (1+R)A_n - x, n = 0, 1, 2, \dots$$

于是可得

$$\begin{aligned} A_n &= A_0(1+R)^n - x[(1+R)^{n-1} + (1+R)^{n-2} + \dots + (1+R) + 1] \\ &= A_0(1+R)^n - x \frac{(1+R)^n - 1}{R}, n = 0, 1, 2, \dots \end{aligned} \quad (1.1)$$

即得 A_n, A_0, x, R, N 之间的关系.

2. 就广告而言

已知 $N = 10$ 年 $= 120$ 个月, $x = 800$ 元, $A_0 = (M - 70\,000)$ 元,则要求 10 年还清,即 $A_{120} = 0$,从而得

$$0 = A_0(1+R)^{120} - \frac{800}{R}[(1+R)^{120} - 1],$$

于是

$$A_0 = \frac{800[(1+R)^{120} - 1]}{R(1+R)^{120}}. \quad (1.2)$$

不妨设月利率 $R=0.01$, 则由 (1.2) 式可算出 $A_0 \approx 55\,760$ 元, 于是房子总价为 $M \approx 70\,000 + 55\,760 = 125\,760$ 元. 由此可知, 一次性付款额不应大于 M , 否则, 就应该自己去银行贷款, 不要借公司的钱了.

3. 推广问题(一)

某高校青年教师张某为买房向公司借贷 $A_0 = 60\,000$ 元, 月利率 $R=0.01$, 若要每月还一次钱, 需 25 年 = 300 个月还清, 张老师希望知道平均每月还多少钱?

根据前面的讨论, 要 25 年 = 300 个月还清贷款, 即要

$$A_{300} = A_0(1+R)^{300} - x \frac{(1+R)^{300} - 1}{R} = 0$$

可以求得 $x \approx 632$ 元, 即平均每月还 632 元, 25 年可还清.

4. 推广问题(二)

如果张老师每半月还一次钱, 每次还 $x = \frac{632}{2} = 316$ 元, 半月利率为 $R = \frac{0.01}{2} = 0.005$, 则能让张老师提前三年(即 22 年)还清, 不过公司要求一次先付三个月的款: $632 \times 3 = 1\,896$ 元作为手续费, 问这种方案对谁有利?

实际上, 表面上看这种方案, 张老师在每月不多还钱的条件下提前三年还清, 对张老师十分有利, 而公司没有多赚钱. 但稍作分析可知, 由于张老师先预付了 1 896 元, 则事实上相当于张老师只借了 $A_0 = 60\,000 - 1\,896 = 58\,104$ 元, 而 $R = 0.005$, $x = 316$ 代入 (1.1) 式, 并令 $A_n = 0$, 则有

$$n = \frac{\ln\left(\frac{x}{x - A_0 R}\right)}{\ln(1 + R)},$$

可解得 $n = 505$ (半月) ≈ 21.04 年, 即提前 3.96 年就还清了借款, 即该公司至少从中多赚了 $632 \times 11.52 = 7\,280.64$ 元.

思考题: 对固定的月利率 R , 如果张老师某一时候想一次付清借款需还多少钱?

1.3.3 儿童保险问题

0 至 17 岁的儿童都可以参加这种保险, 投保金额可以趸交, 也可以按年交, 每份保险金额为 1 000 元, 保险公司要求各年龄的儿童需交投保金额如表 1-1.

表 1-1

| | | | | | | | | | |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 投保年龄 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 年 交 | 599 | 652 | 714 | 787 | 872 | 973 | 1 094 | 1 242 | 1 423 |
| 趸 交 | 5 978 | 6 297 | 6 649 | 7 033 | 7 449 | 7 896 | 8 377 | 8 892 | 9 445 |
| 投保年龄 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 年 交 | 1 605 | 1 888 | 2 266 | 2 795 | 3 584 | 4 886 | — | — | — |
| 趸 交 | 10 036 | 10 669 | 11 346 | 12 070 | 12 843 | 13 669 | 14 551 | 15 492 | 16 496 |

保险公司应对保险人的保险项目和金额为

(1) 教育保险金: 被保险人到 18、19、20、21 周岁时每年可领取一份保险金 1 000 元.

(2) 创业保险金: 被保险人到 22 周岁时可以领取保险金额的 4.7 倍的创业保险金.

(3) 结婚保险金: 被保险人到 25 周岁时可以领取保险金额的 5.7 倍的结婚保险金.

(4) 养老保险金: 被保险人到 60 周岁时可以领取保险金额的 60 倍的养老保险金.

现在的问题是: 如果被保险人能活到 60 岁时, 则

(1) 如果按存款年利率 4.5% 计算, 投保是否合算?

(2) 如果按贷款年利率 8% 计算, 保险公司从中获利多少?

首先假设投保人都能活到 60 岁; 投保人的交款和保险公司的返回保险金均在年初进行; 银行现行的存、贷款利率不变; 这里均按一份投保金额 (1 000 元) 计算.

记投保年龄为 k ($k=0, 1, 2, \dots, 17$); 按年交款额为 a_k , 总交款额为 $A_k = (18-k)a_k$ ($k=0, 1, 2, \dots, 14$); 趸交款额为 B_k ($k=0, 1, 2, \dots, 17$); 银行长期存、贷款利率分别为 $R_1 = 4.5\%$, $R_2 = 8\%$.

1. 问题(一)

如果投保人按年交, 一直交到 17 周岁, 各年龄的投保人共需交款额为

$$A_k = (18-k)a_k \quad (k=0, 1, 2, \dots, 14), \quad (1.3)$$

到 60 周岁时共得保险金额 74 400 元.

设想投保人将所投保的金额存到银行, 到 60 周岁时的存款数额为

$$\begin{aligned}
 X_k &= a_k \sum_{i=k}^{17} (1+R_1)^{i-k} [1 + (1+R_1)^{43}] \\
 &= a_k \frac{(1+R_1)^{18-k} - 1}{R_1} [1 + (1+R_1)^{43}] \quad (k=0, 1, 2, \dots, 14). \quad (1.4)
 \end{aligned}$$

如果投保人按趸交, k 周岁的人交款额为 $B_k (k=0, 1, 2, \dots, 17)$. 设想投保人将所投保的金额存到银行, 到 60 周岁时的存款数额为

$$Y_k = B_k(1 + R_1)^{60-k} \quad (k = 0, 1, 2, \dots, 17). \quad (1.5)$$

由(1.3)、(1.4)、(1.5)式计算结果如表 1-2.

表 1-2

| k | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-------|--------|---------|---------|---------|---------|---------|---------|---------|---------|
| a_k | 599 | 652 | 714 | 787 | 872 | 973 | 1 094 | 1 242 | 1 423 |
| A_k | 10 782 | 11 084 | 11 424 | 11 805 | 12 208 | 12 649 | 13 128 | 13 662 | 14 230 |
| X_k | 98 370 | 98 320 | 98 500 | 98 890 | 99 320 | 99 870 | 100 510 | 101 300 | 102 030 |
| Y_k | 83 860 | 84 530 | 85 410 | 86 450 | 87 620 | 88 880 | 90 230 | 91 660 | 93 160 |
| k | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| a_k | 1 605 | 1 888 | 2 266 | 2 795 | 3 584 | 4 886 | — | — | — |
| A_k | 14 445 | 15 104 | 15 862 | 16 770 | 17 920 | 19 544 | — | — | — |
| X_k | 99 930 | 100 490 | 101 020 | 101 490 | 101 770 | 101 740 | — | — | — |
| Y_k | 94 730 | 96 370 | 98 070 | 99 840 | 101 660 | 103 530 | 105 470 | 107 450 | 109 490 |

从计算结果来看, 显然无论哪个年龄的人是按年交, 还是趸交都不如将钱存到银行合算, 即存款收益都高于投保所得的保险金额.

2. 问题(二)

设投保人按年交, $k (k=0, 1, 2, \dots, 14)$ 周岁的投保人到 18、19、20、21、22、25、60 周岁时公司的收益额分别为

$$Z_k(18) = a_k \sum_{i=k}^{17} (1 + R_2)^{i-k+1} - 1\,000 \quad (k = 0, 1, 2, \dots, 14),$$

$$Z_k(19) = Z_k(18)(1 + R_2) - 1\,000 \quad (k = 0, 1, 2, \dots, 14),$$

$$Z_k(20) = Z_k(19)(1 + R_2) - 1\,000 \quad (k = 0, 1, 2, \dots, 14),$$

$$Z_k(21) = Z_k(20)(1 + R_2) - 1\,000 \quad (k = 0, 1, 2, \dots, 14),$$

$$Z_k(22) = Z_k(21)(1 + R_2) - 4\,700 \quad (k = 0, 1, 2, \dots, 14),$$

$$Z_k(25) = Z_k(22)(1 + R_2)^3 - 5\,700 \quad (k = 0, 1, 2, \dots, 14),$$

$$Z_k(60) = Z_k(25)(1 + R_2)^{35} - 60\,000 \quad (k = 0, 1, 2, \dots, 14).$$

计算可以得到结果如表 1-3.

表 1-3

| k | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-----------|---------|---------|---------|---------|---------|---------|---------|---------|--------|
| $Z_k(60)$ | 174 640 | 162 020 | 150 700 | 140 500 | 130 410 | 120 770 | 111 260 | 101 990 | 91 760 |
| k | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| $Z_k(60)$ | 69 460 | 56 580 | 41 350 | 22 210 | -4 040 | -44 330 | | | |

设投保人按趸交, $k(k=0,1,2,\dots,14)$ 周岁的投保人到 18、19、20、21、22、25、60 周岁时公司的收益额分别为

$$Z_k(18) = B_k(1 + R_2)^{18-k} - 1\,000 \quad (k = 0, 1, 2, \dots, 17),$$

$$Z_k(19) = Z_k(18)(1 + R_2) - 1\,000 \quad (k = 0, 1, 2, \dots, 17),$$

$$Z_k(20) = Z_k(19)(1 + R_2) - 1\,000 \quad (k = 0, 1, 2, \dots, 17),$$

$$Z_k(21) = Z_k(20)(1 + R_2) - 1\,000 \quad (k = 0, 1, 2, \dots, 17),$$

$$Z_k(22) = Z_k(21)(1 + R_2) - 4\,700 \quad (k = 0, 1, 2, \dots, 17),$$

$$Z_k(25) = Z_k(22)(1 + R_2)^3 - 5\,700 \quad (k = 0, 1, 2, \dots, 17),$$

$$Z_k(60) = Z_k(25)(1 + R_2)^{35} - 60\,000 \quad (k = 0, 1, 2, \dots, 17).$$

计算可以得到结果如表 1-4.

表 1-4

| k | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|-----------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| $Z_k(60)$ | 282 860 | 267 930 | 254 750 | 242 860 | 231 950 | 221 690 | 212 070 | 202 900 | 194 240 |
| k | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| $Z_k(60)$ | 185 900 | 177 940 | 170 270 | 162 880 | 155 710 | 148 770 | 142 020 | 135 420 | 128 980 |

从计算结果可以看出,保险公司的收益都是很可观的,特别是趸交的收益高于按年交的收益,但注意到,对于 13 和 14 周岁的投保人按年交公司会赔钱,这也说明 14 周岁以后不能按年交的理由,也反映出这个方案制定的合理性问题.

另外,实际中保险公司还承担着投保人的意外人身保险问题,所以在公司的收益中包含着这部分风险费用,在这里我们没有考虑这部分费用.

思考题: 对保险人采用何种交款方式(趸交、年交)合算?

1.3.4 应急设施的位置问题(AUMCM 1986 - B)

1. 问题的提出

美国的里奥兰翘(Rio Rancho)镇迄今还没有自己的应急设施. 1986 年该镇得到了建立两个应急设施拨款, 每个设施都把救护站、消防队和警察局合在一起. 如图 1-1 指出了 1985 年每个长方形街区发生应急事件的次数, 在北边的 L 形区域是一障碍, 而在南边的长方形区域内是一个有浅水塘的公园. 应急车辆驶过一条南北向的街道平均要花 15 s, 而通过一条东西向的街道平均花 20 s. 你的任务是确定这两个应急设施的位置, 使得总响应时间最少.

(I) 假定应急需求集中在每个街区的中心, 而应急设施位于街角处;

(II) 假定应急需求是沿包围每个街区的街道上均匀分布, 而应急设施可以位于街道的任何地方.

2. 模型的假设

- (1) 两个障碍区域中均不需要应急服务;
- (2) 每年的应急事件数目比较小, 可以认为在同一街区不会同时发生两个事件;
- (3) 忽略车辆拐弯和过十字路口的时间, 仅考虑沿街道行驶的时间;
- (4) 两个设施的功能相同, 当应急事件发生时, 指挥中心总是从离事件发生地最近的应急设施派出应急车辆;
- (5) 1985 年的各街区的应急事件数是真实的, 未来的需求分布不会与现在的需求相差太远;
- (6) 当连接两点的不同路径所用时间相同时, 路径可以任选其一.

3. 模型的建立与求解

模型 I:

除了上面假设以外, 假设在没有障碍的街区应急事件均发生在街区中心, 而应急设施的位置设在某街区的街角上. 应急车辆做出响应的最短时间是指到达事件发生点的时间; 这样可能的两个应急设施的位置点数只有有限个, 因此, 只需要检验每一对位置点对所有街区发生事件做出的响应时间, 选择平均每一次事件响应时间最小的那两个点建立坐标系, 左下角(西南角)为原点(0,0), 东西为 x 轴, 南北为 y 轴.

(1) 一个位置点对某一街区发生事件的响应时间 = 位置点到街区的街道数

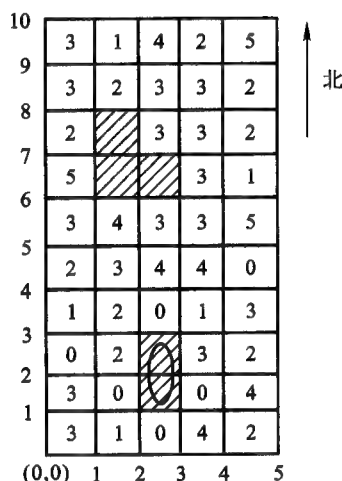


图 1-1

× 车辆行驶一条街道的时间 × 该街区发生事件的次数；

(2) 一个位置点对全镇所有应急事件的响应时间总和 = 该位置点对所有街区的应急事件响应时间的总和；

(3) 一个位置点对全镇任一次应急事件的平均响应时间 = 总响应时间/事件的总数；

(4) 取使平均响应时间最小的那个对应的位置点为应急设施的位置。

两个设施到任一街区 (X, Y) (距原点最近的街角坐标) 的时间计算公式为

$$T1 = |X1 - (X + 0.5)| \times 20 + |Y1 - (Y + 0.5)| \times 15 - 17.5,$$

$$T2 = |X2 - (X + 0.5)| \times 20 + |Y2 - (Y + 0.5)| \times 15 - 17.5,$$

其中 $(X1, Y1)$ 表示第一设施的位置坐标, $(X2, Y2)$ 表示第二设施的位置坐标; 0.5 是因为 X, Y 分别表示街区左下 (西南) 角的坐标, $(X + 0.5, Y + 0.5)$ 是表示街区中心的坐标, 设施到街区的距离为设施到街区中心的距离. “-17.5” 是因车辆穿过一条东西街道要用 20 s, 南北用 15 s, 前面的距离算到了街区中心, 而车辆行驶只到最邻近的街角上, 因此东西减去 10 s, 南北减去 7.5 s. 取最邻近的一个设施所需时间:

$$TM = \min(T1, T2).$$

由以 (X, Y) 为坐标的街区发生事件的次数 $W(X, Y)$, 可以求出两个设施到任意街区最邻近的街角所需的时间:

$$TOT = TM \times W(X, Y),$$

求总响应时间: $T = \sum TOT$,

平均响应时间: $T/109$ s.

经计算可得, 两个应急设施的位置分别为 $(3, 4)$ 和 $(3, 8)$, 并且可算出从这两个设施到任意一个街区最邻近的街角上的平均时间为 29.5 s (如图 1-2). 这是最佳的两个位置, 其他的任何地方的相应时间都会大于 29.5 s. 还注意到从这两个位置到邻近障碍区的街区并不因为障碍增加时间。

模型 II:

除了前面的假设以外, 假设每个街区的应急事件都发生在该街区四周的街道上, 而且均匀分布, 两个设施还是设在街角上。

基本上采用模型 I 的方法过程, 注意到由于可能的事件发生点在街道上均匀分布, 为此, 在每一条街道上的事件发生点不必一点点的考虑, 可以认为每一条街道上发生的事件都集中在一点上 (类似于均匀分布密度的直线质量可以认为集中在一点上——质心), 该点应该是从这一点到街角的距离等于到实际事件发生点的平均距离, 这一点一定是在街道的中心. “每一个方形街区四周的每

一条街道上发生事件的次数 = 该方形街区事件数的 $\frac{1}{4}$ ”。因此，“每一条街道上发生事件的次数 = 两个相邻街区事件数的 $\frac{1}{4}$ 之和”。

首先求两个设施的位置到任意街道中心(事件点)处所需要的时间:

$$T1 = |X1 - X| \times 20 + |Y1 - Y| \times 15,$$

$$T2 = |X2 - X| \times 20 + |Y2 - Y| \times 15.$$

注: 这里因应急车辆不需要到街区中心, 也不需要过街道, 所以, 不需要减 17.5, 和加 0.5。

由 $TM = \min(T1, T2)$, 再求两个设施到任意街道中点所需要的总时间, 分为两个部分:

东西方向的为

$$TOT = TM \times \frac{1}{4} [W(X - 0.5, Y) + W(X - 0.5, Y - 1)],$$

南北方向的为

$$TOT = TM \times \frac{1}{4} [W(X, Y - 0.5) + W(X - 1, Y - 0.5)],$$

最后求总的响应时间: $T = \sum TOT$ 。

平均响应时间为 $T/109$ s。

经计算可得两个设施的最合适的位置是 (3,4) 和 (3,8), 同模型 I, 平均响应时间为 47.0 秒。从响应时间来看模型 II 比模型 I 多了 17.5 秒, 这是由模型 II 的条件所决定的。主要是模型 II 的车辆可行驶到实际事件点, 而模型 I 中的车辆只能到发生事件的街区最邻近的角上(如图 1-3)。

该模型从分布图上可以看出, 车辆到绕过障碍区的街道上去也不会增加时间。

4. 模型的分析

(1) 因为问题中给出的仅是 1985 年一年

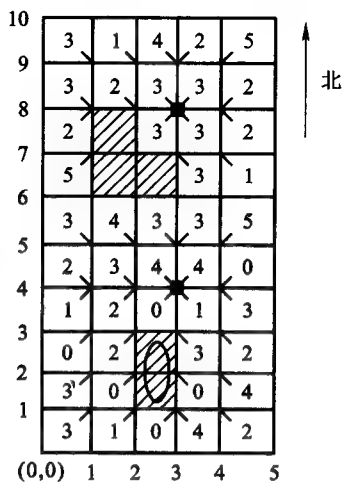


图 1-2

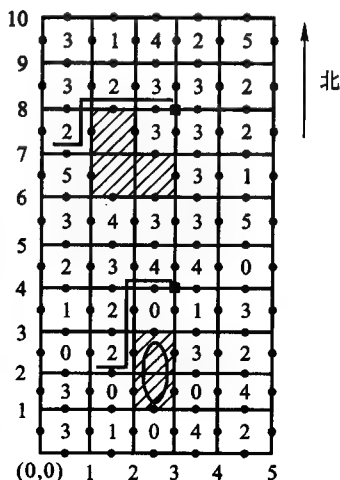


图 1-3

的事件分布数据,模型中不可能做更多的计算,或者作图等,如果所给数据覆盖几年的话就好了,模型中对于障碍没有充分地考虑,当然如果给出环绕障碍物弯曲的有关数据,可以做得更好,但模型会复杂.

(2) 模型中全部忽略了车辆转弯的时间,这种假设不会有太大的影响,对模型 I 的任何路线至多有一个转弯,对模型 II 只有两条路线有两个转弯,其他至多有一个转弯.

(3) 在前面的假设中,设施的应急车辆只被派往正常范围内的应急事件点,即便是被派往正常范围以外,影响也不太大.

(4) 假设中把应急设施设在街道交叉口处,可对任何方向的应急事件灵活地做出响应,指挥中心可以随机应变地调动车辆,更容易转弯、调头等.

1.3.5 几个简单的数学建模问题

问题(1): 储蓄问题

张老师现有本金 N 元,打算留给孩子 10 年以后上大学使用,他想以“整存整取”的方式存入银行,请你根据银行现行的“整存整取”利率表(分三个月、半年、一年、二年、三年、五年、八年),选择一种合适的存款方式组合,使 10 年后获息最多.

问题(2): 贷款修桥问题

某市政府拟贷款 10 亿元人民币修建一条高速公路,年利率 8%,按设计方案预言公路建成后每天收车辆过路费 35 万元.另外,每年养路费和职工工资等开支费用为 2 000 万元.问

(1) 该市政府需要多少年才能还清这笔贷款?

(2) 如果每天所收车辆过路费只有 30 万元,那么该市政府能否还清贷款?

(3) 如果银行要求必须在 15 年内还清贷款,那么每天收过路费至少要多多少元?

问题(3): 军备竞赛问题

目前,世界并不太平,国与国之间和地区之间的种族歧视、民族矛盾、经济利益冲突等问题造成的局部战争和地区的武装冲突时有发生,有的长期处于敌对状态,从而导致了军备竞赛.随着高科技的发展,军事装备现已成为决定战争胜负的重要因素.在这里我们所说的军备是指军事实力的总和,主要包括军事装备、军事兵力、军事费用等.实际中,引起军备竞赛的因素很多,试分析与军备竞赛有关的诸多因素,建立相应的数学模型,并就相互之间的关系进行讨论.

问题(4): 选择评委问题

某电视台准备举行青年歌手大奖赛,拟从 12 名评委中选择 10 名高水平的评委,主要是依据上一年这 12 名评委为 10 名歌手打分的情况来选择,具体数据

如表 1-5.

表 1-5

| 歌手 \ 评委 | | | | | | | | | | | | |
|---------|------|------|------|------|------|------|------|------|------|------|------|------|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | 9.10 | 9.10 | 9.12 | 9.14 | 9.17 | 9.18 | 9.20 | 9.21 | 9.23 | 9.25 | 9.28 | 9.30 |
| 2 | 9.00 | 8.98 | 8.95 | 9.13 | 9.15 | 9.16 | 9.00 | 9.20 | 9.23 | 9.26 | 9.28 | 9.31 |
| 3 | 8.89 | 8.83 | 8.90 | 9.10 | 9.22 | 9.16 | 9.18 | 9.25 | 9.22 | 9.30 | 9.30 | 9.32 |
| 4 | 9.20 | 9.00 | 9.10 | 9.21 | 9.23 | 9.30 | 9.23 | 9.20 | 9.21 | 9.25 | 9.26 | 9.28 |
| 5 | 8.98 | 8.88 | 9.00 | 9.13 | 9.21 | 9.24 | 9.26 | 9.23 | 9.27 | 9.28 | 9.20 | 9.30 |
| 6 | 9.13 | 9.11 | 9.14 | 9.16 | 9.14 | 9.16 | 9.20 | 9.26 | 9.23 | 9.26 | 9.28 | 9.31 |
| 7 | 9.23 | 9.10 | 9.24 | 9.18 | 9.25 | 9.23 | 9.27 | 9.29 | 9.26 | 9.27 | 9.26 | 9.36 |
| 8 | 9.12 | 9.03 | 9.25 | 9.25 | 9.21 | 9.23 | 9.28 | 9.30 | 9.28 | 9.26 | 9.30 | 9.40 |
| 9 | 9.30 | 9.20 | 9.30 | 9.34 | 9.32 | 9.35 | 9.29 | 9.32 | 9.34 | 9.36 | 9.32 | 9.45 |
| 10 | 9.20 | 9.12 | 9.25 | 9.30 | 9.28 | 9.33 | 9.23 | 9.30 | 9.34 | 9.27 | 9.38 | 9.40 |

请你对评委的评判水平做出评价,并确定出 10 名高水平的评委参加今年的评判工作.

1.4 数学建模的方法和步骤

1.4.1 数学建模方法

数学建模一般是通过问题的实际背景,给出一些已知信息,这些信息可以是一组实测数据或模拟数据,也可以是若干参数、图形,或者仅给出一些定性描述,依据这些信息建立数学模型的方法有很多,但从基本解法上可以分为五大类:

(1) 机理分析方法:主要是根据实际中的客观事实进行推理分析,用已知数据确定模型的参数,或直接用已知参数进行计算.

(2) 构造分析方法:首先建立一个合理的模型结构,再利用已知信息确定模型的参数,或对模型进行模拟计算.

(3) 直观分析方法:通过对直观图形、数据进行分析,对参数进行估计、计算,并对结果进行模拟.

(4) 数值分析方法:对已知数据进行数值拟合,可选用插值方法、差分方法、样条函数方法、回归分析方法等.

(5) 数学分析方法: 用“现成”的数学方法建立模型, 如图论、微分方程、规划论、概率统计方法等.

在实际建模的过程中, 根据问题的实际背景和已知信息选择适当方法, 尽量使用“现成”的数学方法. 如果已知信息不明确, 或不完整时, 可以进行适当补充或舍弃, 甚至可以修改题目的条件、参数和数据. 也可以先做最简单的模型, 然后再逐步地完善改进.

数学建模或参加建模竞赛一般应具备的方法和知识: 一是要掌握常用的建模方法, 如机理分析法、层次分析法、差分法、图论法、插值与拟合法、统计分析法、优化方法等. 二是要有广泛的知识, 特别是必备的数学知识, 如微分方程、概率统计、规划论、图与网络、数值计算、排队论、对策论、决策论等. 另外, 还应了解一些现代应用数学的知识, 如模糊数学、灰色理论、时间序列、神经网络等. 这些都是数学建模教学的内容, 数学建模所需要的知识首先是“广”, 其次才是“精”. 同时, 在教学中还应介绍一些典型的数学模型案例, 以及实际中政治、经济、工业、农业、商业管理、日常生活中的建模实例等内容.

1.4.2 数学建模的步骤

数学建模是一种创造性的过程, 它需要相当高的观察力、想像力和灵感. 数学建模的过程是有一定的阶段性的, 要解决的问题都是来自于现实世界之中. 数学建模的过程就是对问题进行分析、提炼, 用数学语言做出描述, 用数学方法分析、研究、解决, 最后回到实际中去, 应用于解决和解释实际问题, 乃至更进一步地作为一般模型来解决更广泛的问题. 数学建模的流程为

实际问题→抽象、简化问题, 明确变量和参数→根据某种定律建立变量和参数间数学关系(数学模型)→解析地或近似地求解该数学模型→解释、验证求解结果→应用于实际.

对我们来说, 这一过程为

问题分析→模型假设→模型建立→模型求解→解的分析与检验→论文写作→应用实际.

(1) 问题的分析

数学建模的问题, 通常都是来自于实际中的各个领域的实际问题, 没有固定的方法和标准的答案, 因而既不可能明确给出该用什么方法, 也不会给出恰到好处的条件, 有些时候所给出的问题本身就是含糊不清的. 因此, 数学建模的第一步就应该是对问题所给的条件和数据进行分析, 明确要解决的问题. 通过对问题的分析, 明确问题中所给出的信息、要完成的任务和所要做的、可能用到的知识和方法、问题的特点和限制条件、重点和难点、开展工作的程序和步骤等. 同时, 还要明确题目所给条件和数据在解决问题中的意义和作用、本质的和非本质

的、必要的和非必要的等等. 从而, 可以在建模的过程中, 适当地对已有的条件和数据进行必要的简化或修改, 也可以适当地补充一些必要的条件和数据.

(2) 模型的假设

实际中, 根据问题的实际意义, 在明确建模目的的基础上, 对所研究的问题进行必要的、合理的简化, 用准确简练的语言给出表述, 即模型的假设, 这是数学建模的重要一步. 合理假设在数学建模中除了起着简化问题的作用外, 还对模型的求解方法和使用范围起着限定作用. 模型假设的合理性问题是评价一个模型优劣的重要条件之一, 也是模型的建立成败的关键所在, 假设做的过于简单, 或过于详细, 都会可能使得模型建立的不成功. 为此, 实际中要做出合适的假设, 需要一定的经验和探索, 有时候需要在建模的过程中对已做的假设进行不断地补充和修改.

(3) 模型的建立

在建立模型之前, 首先要明确建模的目的, 因为对于同一个实际问题, 出于不同的目的所建立的数学模型可能会有所不同. 在通常情况下, 建模的目的可以是描述或解释现实世界的现象; 也可以是为了预报一个事件是否会发生, 或未来的发展趋势; 也可以是为了优化管理、决策或控制等. 如果是为了描述或解释现实世界, 则一般可采用机理分析的方法去研究事物的内在规律; 如果是为了预测预报, 则常常可以采用概率统计、优化理论或模拟计算等有关的建模方法; 如果是为了优化管理、决策或控制等目的, 则除了有效地利用上述方法之外, 还需要合理地引入一些量化的评价指标以及评价方法. 对于实际中的一个复杂的问题, 往往是要综合运用多种不同方法和不同学科的知识来建立数学模型, 才能够很好地解决这一个问题. 在明确建模目的的基础上, 在合理的假设之下, 就可以完成建立模型的任务, 这是我们数学建模工作中最重要的一个环节. 根据所给的条件和数据, 建立起问题中相关变量或因素之间的数学规律, 可以是数学表达式、图形和表格, 或者是一个算法等, 都是数学模型的表示形式, 这些形式有时可以相互转换.

(4) 模型的求解

不同的数学模型的求解方法一般是不同的, 通常涉及不同数学分支的专门知识和方法, 这就要求我们除了熟练地掌握一些数学知识和方法外, 还应具备在必要时针对实际问题学习新知识的能力. 同时, 还应具备熟练的计算机操作能力, 熟练掌握一门编程语言和一两个数学工具软件包的使用. 不同的数学模型求解的难易程度是不同的. 一般情况下, 对较简单的问题, 应力求普遍性; 对较复杂的问题, 可从特殊到一般的求解思路来完成.

(5) 解的分析与检验

对于所求出的解, 必须要对解的实际意义进行分析, 即模型的解在实际问题

中说明了什么、效果怎样、模型的适用范围如何等等.同时,还要进行必要的误差分析和灵敏度分析等工作.由于数学模型是在一定的假设下建立的,而且利用计算机的近似求解,其结果产生一定的误差是必然的.通常意义下的误差主要来自于由模型的假设引起的误差、近似求解方法产生的误差、计算机产生的舍入误差和问题的数据本身误差.实际中,对这些误差很难准确地给出定量估计,往往是针对某些主要的参数做相应的灵敏度分析,即当一个参数有很小的扰动时,对结果的影响是否也很小,由此可以确定相应变量和参数的误差允许范围.

(6) 论文写作

因为数学建模工作的目的是为了解决实际问题,所以工作完成以后要写出一篇论文,即等于一篇研究报告.论文要力图通俗易懂,能让人明白你用什么方法解决了什么问题,结果如何,有什么特点.为此,应尽可能使论文的表述清晰、主题明确、论述严密、层次分明、重点突出、符合科技论文的写作规范.同时,要注意论文的写作工作是贯穿始终的,在建模的每个阶段都应该把你的主要思路和工作写下来,这是论文写作时的第一手材料.

(7) 应用实际

对于所建立的数学模型以及求解结果,只有拿到实际中去应用检验后,才被证明是正确的.否则,就需要修正模型的假设或条件,重新建立模型,直到通过实际的检验为止,方可应用于实际.

1.5 数学建模与能力培养

通过学习数学建模这门课程和从事这方面的工作,主要是扩大学生的知识面,培养和提高学生综合运用所学知识解决实际问题的综合能力,即“数学建模的能力”.具体地讲,数学建模有利于培养以下几个方面的能力:

(1) 丰富灵活的想像能力:数学建模要解决的问题往往都需要多学科的知识 and 多种不同的方法,因此,需要我们具备丰富的想像能力,有人说:“想像力是最高的天赋——是一种把原始经历组合成具体形象的能力,一种把握层次能力,一种把感觉、梦幻和理想等对立因素融合成一个统一整体的能力.”

(2) 抽象思维的简化能力:实际中的问题往往都是很复杂的,数学建模的过程就是通过对问题进行抽象、简化将其转化为数学问题.因此,这种抽象思维的简化能力是必不可少的,数学建模的学习和训练有利于培养这种能力.

(3) 一眼看穿的洞穿能力:洞察能力是一种直觉的领悟,是把握事物内在的或隐藏的和本质的能力,简言之就是“一眼看穿”的能力.这种能力对于数学建模是非常重要的,但需要经过艰苦的、长期的经验积累和有针对性地训练.

(4) 发散思维的联想能力:发散思维是发明创造的一个有力武器,在数学

建模的过程中,通过某些关键信息展开联想,这是一种“由此及彼,由彼及此”的能力。

(5) 与时俱进的开拓能力:随着社会的进步和发展,科学技术也快速地进行,实际中的问题复杂多变,数学建模也必须与时俱进,发扬开拓精神,培养创新能力,这也是新型创新人才素质的一部分。

(6) 学以致用应用能力:学以致用是21世纪高素质应用型人才所具备的一种素质,因为一个人所掌握的知识总是有限的,但解决实际问题所需要的知识相对是无限的,因此,我们必须具备这种急用先学,学以致用应用能力,数学建模是培养我们这种能力的一种有效途径。

(7) 会抓重点的判断能力:数学建模的问题所给条件和数据往往不是恰到好处,有时也可能是杂乱无章的,这就要求我们具备特有的一种会抓重点的判断能力,充分利用已知信息,寻找突破口,来解决问题。

(8) 高度灵活的综合能力:因为数学建模的问题是综合性的,解决问题所需要的知识和方法也是综合性的,因此,我们能力也必须是综合性。否则,我们将是“只见树木,不见森林”,不可能完整地解决问题。

(9) 使用计算机的动手能力:数学建模必须要熟练掌握计算机的操作,以及工具软件的使用和计算编程,这是因为对实际问题进行分析和建立数学模型以后的求解都有大量的推理运算、数值计算、作图等工作,这都需要通过计算机和软件技术来实现。

(10) 信息资料的查阅能力:信息资料的查阅能力是科技人才所必备、数学建模所必需的能力。

(11) 科技论文的写作能力:论文的写作能力是数学建模的基本技能之一,也是科技人才的基本能力之一,是表达我们所做工作的惟一方式。通过论文,要让读者清楚地知道用什么方法解决了什么问题,结果为何,效果怎么样等等。

(12) 团结协作的攻关能力:数学建模都是以小组为单位开展工作的,体现的是团队精神,培养的是团结协作的能力,也是未来科研工作所必备的能力,不具备这种能力的人则将一事无成。

1.6 参考案例与参考文献

1. 参考案例

- (1) 人口增长问题——文献[1]:13—18
- (2) 公平席位的分配问题——文献[1]:29—34
- (3) 卫星发射问题——文献[2]:7—10
- (4) 疏散问题——文献[3]:8—10

(5) 渡口问题——文献[3]:15—19

(6) 公共汽车问题——文献[4]:85—90

2. 参考文献

[1] 姜启源. 数学模型. 第二版. 北京:高等教育出版社,1993

[2] 陈义华. 数学模型. 重庆:重庆大学出版社,1995

[3] 徐全智,杨晋浩. 数学建模入门. 成都:电子科技大学出版社,1996

[4] 徐全智,杨晋浩. 数学建模. 北京:高等教育出版社,2003

[5] 叶其孝. 大学生数学建模竞赛辅导教材(二). 长沙:湖南教育出版社,2001

[6] 李大潜. 中国大学生数学建模竞赛. 第二版. 北京:高等教育出版社,2001

[7] 张思明. 中学数学建模教学的实践与探索. 北京:北京教育出版社,1998

[8] 韩中庚. 浅谈数学建模与人才培养. 工程数学学报. 2003, 20(5): 119—123

第二章 两种初等分析方法

所谓的初等分析方法主要是指所用的数学知识和方法都是初等的,而不是高等的.在解决实际问题的过程中,往往主要是看解决问题的效果和应用的结果如何,而不在于用了初等的方法还是高等的方法.对于数学建模也是这样,判断一个数学模型的优劣完全在于模型的正确性和应用效果,而不在于采用多少高深的数学知识.然而,在同样的应用效果之下,用初等方法建立的数学模型可能更优于用高等方法建立的数学模型.通常我们所说的初等分析建模方法有很多,常用的像类比分析方法、几何分析方法、逻辑分析方法、量纲分析方法、集合分析方法等.这些方法主要是根据对现实对象特性的认识,分析其因果关系,找出反映内部机理的规律.所建立的数学模型一般有明确的物理意义或现实意义,对现实对象的认识主要来源于两个方面:一是与问题相关的物理、化学、经济等自然学科方面的知识;二是通过对数据和现象的分析对事物的内在规律做出合理的假设.因此,这些简单的建模方法也都属于机理分析方法的范畴.在这里仅介绍量纲分析方法与集合分析方法及应用.

2.1 量纲分析方法

量纲分析是物理学中常用的一种定性分析方法,也是在物理领域中建立数学模型的一个有力工具.利用这种方法可以从某些条件出发,对某一物理现象进行推断,可将这个物理现象表示为某些具有量纲的变量的方程,从而可以用此来分析各物理量之间的关系.

2.1.1 量纲

当我们对一个物理概念进行定量描述时,总离不开它的一些特性,比如,时间、质量、密度、速度、力等等,这种表示不同物理特性的量,称之为具有不同的“量纲”,记为 $[\bullet]$.另外,物理量的度量又离不开单位,比如质量的单位可以为 g, kg ,速度的单位为 $cm/s, m/s, km/h$ 等.但当用数学公式描述一物理量时,等号的两端就必须保持量纲的一致性和单位的一致性.换言之,量纲分析就是基于量纲一致的原则来分析物理量之间关系的一种方法.

实际中,有些物理量的量纲是基本的,称为**基本量纲**,而其他的物理量的量纲可由这些基本量纲来表示.例如,在力学中,基本量纲是质量的量纲 $[M]$ 、长度的量纲 $[L]$ 和时间的量纲 $[T]$.而速度的量纲为 $[V] = [LT^{-1}]$,加速度的量纲为 $[a] = [LT^{-2}]$.根据 Newton 第二定律: $F = ma$,则力的量纲为 $[F] = [M][LT^{-2}] = [MLT^{-2}]$.实际中,也有些量是无量纲的,比如常数 π, e 等,此时记为 $[\pi] = [e] = 1$.

2.1.2 量纲齐次原则

当用数学公式表示一个物理定律时,等号两端必须保持量纲的一致性,这种性质称为**量纲齐次性**.当方程中的各项具有相同的量纲时,这个方程被称为是**量纲齐次的**,也只有具有相同量纲的量才能作比较或相加、减,由此可知,物理定律必须是量纲齐次的.例如,万有引力定律: $f = k \frac{m_1 m_2}{r^2}$,其引力常数 $k = \frac{fr^2}{m_1 m_2}$ 的量纲为

$$[k] = [MLT^{-2}] \cdot [M^{-2}L^2] = [M^{-1}L^3T^{-2}],$$

所以 k 是一个有量纲的常数.又例如动量定律: $mv_2 - mv_1 = f\Delta t$,左边的量纲为 $[MLT^{-1}]$,右边的量纲为 $[F][T] = [MLT^{-2}][T] = [MLT^{-1}]$,即量纲是齐次的.

根据量纲齐次性原则,可以有下面的量纲分析法的基本定理.

定理(Buckingham Pi) 设有 m 个物理量 q_1, q_2, \dots, q_m 满足某定律:

$$f(q_1, q_2, \dots, q_m) = 0,$$

X_1, X_2, \dots, X_n 是基本量纲($n \leq m$). q_j 的量纲可以表示为 $[q_j] = \prod_{i=1}^n X_i^{a_{ij}} (j = 1, 2, \dots, m)$. 矩阵 $A = (a_{ij})_{n \times m}$ 称为**量纲矩阵**,若 A 的秩 $\text{rank } A = r$,可设线性齐次方程组 $AY = 0$ (Y 是 m 维向量),有 $m - r$ 个基本解为 $y_k = (y_{k1}, y_{k2}, \dots, y_{km})^T (k = 1, 2, \dots, m - r)$. 则 $\pi_k = \prod_{j=1}^m q_j^{y_{kj}}$ 为 $m - r$ 个相互独立的无量纲的量,且有 $F(\pi_1, \pi_2, \dots, \pi_{m-r}) = 0$ 与 $f(q_1, q_2, \dots, q_m) = 0$ 等价,其中 F 为一未知函数.

2.1.3 量纲分析的一般步骤

(1) 将与问题有关的物理量(变量和常量)收集起来,记为 q_1, q_2, \dots, q_m ,根据问题的物理意义确定基本量纲,记为 $X_1, X_2, \dots, X_n (n \leq m)$;

(2) 写出 q_j 的量纲 $[q_j] = \prod_{i=1}^n X_i^{a_{ij}} (j = 1, 2, \dots, m)$;

(3) 设 q_1, q_2, \dots, q_m 满足关系 $\pi = \prod_{j=1}^m q_j^{y_j}$,其中 y_j 为待定的, π 为无量纲的

量,因此 $[\pi] = \prod_{j=1}^m X_j^{a_j} = 1$, 于是 $a_i = \sum_{j=1}^m a_{ij} y_j = 0 \ (i = 1, 2, \dots, n)$;

(4) 解线性方程组

$$\sum_{j=1}^m a_{ij} y_j = 0 \ (i = 1, 2, \dots, n),$$

系数矩阵 $A = (a_{ij})_{n \times m}$, $\text{rank } A = r$, 则方程组有 $m - r$ 个基本解, 记

$$y_k = (y_{k_1}, y_{k_2}, \dots, y_{k_m})^T, k = 1, 2, \dots, m - r;$$

(5) 记 $\pi_k = \prod_{j=1}^m q_j^{y_{kj}}$, 则 $\pi_k \ (k = 1, 2, \dots, m - r)$ 为无量纲的量;

(6) 由 $F(\pi_1, \pi_2, \dots, \pi_{m-r}) = 0$ 解出物理规律.

2.2 空间点热源的扩散问题

2.2.1 问题的提出

设初始时刻 ($t=0$) 在空间中有一热量为 e 的瞬时热源位于原点处 ($r=0$), 热量通过介质向无穷空间扩散, 试研究点热源的扩散规律.

2.2.2 模型的假设

- (1) 任意时刻 t , 空间任一点 (径向距离为 r) 的温度为 u ;
- (2) 介质的初始温度为 0;
- (3) 问题的基本量纲为长度、质量、时间和温度的量纲, 即 $L, M, T, [u] = \theta$.

2.2.3 模型的建立与求解

由热学知识可知, 在 t 时刻, 空间任一点 (径向距离为 r) 的温度为 $u = u(r, t, e, c, k)$, 其中 c 为介质的体积比热, 即单位体积的介质温度升高一度所需热量. k 为介质的扩散系数, 即由 $q = -k \frac{\partial u}{\partial r}$ 确定 (q 是单位时间通过单位面积的热量).

根据 Buckingham Pi 定理, 可设 $f(u, r, t, e, c, k) = 0$. 由假设则有 $[r] = L, [t] = T, [e] = L^2 MT^{-2}$ 为热量 e 的量纲, 此与功的量纲相同, 即为 $[F] \cdot [r] = L^2 MT^{-2}$ (热量是单位质量在热场的作用下移动一段距离所作的功), $[c] = \frac{[e]}{L^3 [u]} = L^{-1} MT^{-2} \theta^{-1}, [k] = LMT^{-3} \theta^{-1}$.

事实上, 由

$$\left[\frac{\partial u}{\partial r}\right] = \frac{[u]}{L} = \frac{\theta}{L}, [q] = [k] \left[\frac{\partial u}{\partial r}\right] = \frac{[e]}{TL^2},$$

则

$$[k] = \frac{[e]}{TL^2 \left[\frac{\partial u}{\partial r}\right]} = \frac{[e]}{TL^2} \frac{L}{\theta} = \frac{[e]}{TL\theta} = LMT^{-3}\theta^{-1}.$$

于是量纲矩阵为

$$A = \begin{bmatrix} 0 & 1 & 0 & 2 & -1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & -2 & -2 & -3 \\ 1 & 0 & 0 & 0 & -1 & -1 \end{bmatrix} \begin{matrix} L \\ M \\ T \\ \theta \\ u \\ r \\ t \\ e \\ c \\ k \end{matrix},$$

其 $\text{rank } A = 4$, 线性齐次方程组 $AY = 0$ 有 $6 - 4 = 2$ 个基本解, 可取

$$\begin{cases} y_1 = (0, 2, -1, 0, 1, -1)^T, \\ y_2 = (-2, 0, -3, 2, 1, -3)^T. \end{cases}$$

于是可得两个相互独立的无量纲的量

$$\begin{cases} \pi_1 = r^2 t^{-1} c k^{-1}, \\ \pi_2 = u^{-2} t^{-3} e^2 c k^{-3} \end{cases}$$

且有 $F(\pi_1, \pi_2) = 0$ 与 $f(u, r, t, e, c, k) = 0$ 等价.

根据隐函数存在定理可得 $\pi_2 = g(\pi_1)$, 即

$$u = \frac{e}{c} (a^2 t)^{-\frac{3}{2}} g\left(\frac{r^2}{a^2 t}\right),$$

其中 $a^2 = \frac{k}{c}$, g 为待定的.

在此 g 无法确定, 量纲分析法只能给出温度函数依赖于其他参数的关系, 具体的形式不能给出. 因为这些函数是无量纲的, 实际中有时可以通过实验数据进行模拟检验.

另一方面, 该问题可用热传导方程求解

$$u = \frac{e}{c} \left(\frac{1}{2\sqrt{\pi a^2 t}} \right)^3 e^{-\frac{r^2}{4a^2 t}}.$$

注: 基本量纲的选取方法不是惟一的, 此问题也可经选取 $\theta, L, T, [e] = E$ 为基本量纲, 类似地可得到相应的结果.

由上面的实例可以看出: 量纲分析方法虽然可以得到一些重要有用的结

果,但有很大的局限性,主要是用于对实际问题的定性描述.

2.3 集合分析方法

集合是数学的一个基础概念,集合论是研究集合的一般性质的一个数学基础分支,也是数学分析、实变函数论的重要基础. 一般认为,集合与集合论的有关概念和理论都是非常抽象的,不便于在实际中应用. 但事实上,对于实际中的许多问题,往往是其相关因素之间的关系比较复杂,用相对通俗、简单的语言文字难以表达,而用集合的概念、术语和子集合之间的运算关系来解释、描述这个实际问题可能更清晰、更直接、更方便,并借助于集合论的相关理论得到具有实际意义的结果. 这就是我们所说的集合分析方法.

2.3.1 集合的概念

所谓的**集合**是指具有某种特定性质的(具体或抽象)事物全体,其构成集合的每个事物(或成员)称为**集合的元素**. 通常用大写字母 A, B, C, X, Y, Z 等表示集合,而用小写字母 a, b, c, x, y, z 等表示集合的元素.

设 A 表示一个集合,对于一个事物 x 可能是集合 A 的一个元素,也可能不是集合 A 的元素,且两者必属其一. 如果 x 是集合 A 的元素,则称 x 属于集合 A ,记为 $x \in A$,否则称 x 不属于集合 A ,记为 $x \notin A$.

当 A 是具有某种性质 P 的元素全体构成的集合时,则通常表示为

$$A = \{x \mid x \text{ 具有的性质 } P\}.$$

例如:所有参加过 2004 年中国大学生数学建模竞赛(CUMCM)的大学生构成集合 A ,则用 x 表示在校的大学生,集合 A 可以表示为

$$A = \{x \mid x \text{ 参加了 2004 年 CUMCM}\}.$$

如果集合 A 是由实数(或实数组)组成的,则又称 A 为**点集**;如果集合 A 是由向量组成的,则称 A 为**向量集**.

设 A, B 是两个集合,如果 A 中的元素都是 B 中的元素,则称集合 A 是集合 B 的**子集**,记为 $A \subset B$,即 A 包含在 B 中,或记为 $B \supset A$,即 B 包含 A . 显然有 $A \subset A$. 如果 $A \subset B$,且 $B \subset A$,即集合 A 与集合 B 的元素相同,则称集合 A 与集合 B 相等,或集合 A 等于集合 B ,记为 $A = B$.

如果一个集合中不含有任何元素,则称此集合为**空集**,通常记为 \varnothing .

2.3.2 集合的运算

设 A, B 是两个集合,由集合 A 与集合 B 的元素所组成的集合称为 A 与 B 的**并集**(或**和集**),简称为**并**(或**和**),记为 $A \cup B$. 将所有既属于集合 A 又属于集合 B

的元素组成的集合称为 A 与 B 的交集(或通集), 简称为交(或通), 记为 $A \cap B$.

完全类似地, 可以将并与交的运算推广到任意有限个, 或无限个集合的情况, 即有限个集合的并与交为 $\bigcup_{i=1}^N A_i$ 与 $\bigcap_{i=1}^N A_i$; 无限个集合的并与交为 $\bigcup_{i=1}^{\infty} A_i$ 与 $\bigcap_{i=1}^{\infty} A_i$.

由集合 A 中不属于集合 B 的元素全体所组成的集合称为集合 A 与集合 B 的差集, 记为 $A - B$. 特别地, 当 $B \subset A$ 时, 称差集 $A - B$ 为 B 关于 A 的余集, 记为 $C_A B$. 又称 $(A - B) \cup (B - A)$ 为集合 A 与集合 B 的对称差, 记为 $A \Delta B$.

设 A, B 是两个集合, 当 $A \cap B = \emptyset$ 时, 则称集合 A 与 B 不交, 或独立. 当 $A \cap B \neq \emptyset$ 时, 则称集合 A 与集合 B 相交.

由集合的并、交和差的定义, 不难证明集合的运算具有如下的一些性质:

- (1) 并与交的幂等性: $A \cup A = A, A \cap A = A$;
- (2) 空集是并运算的零元: $A \cup \emptyset = A$;
- (3) 并的交换律: $A \cup B = B \cup A$;
- (4) 交的交换律: $A \cap B = B \cap A$;
- (5) 并的结合律: $(A \cup B) \cup C = A \cup (B \cup C)$;
- (6) 交的结合律: $(A \cap B) \cap C = A \cap (B \cap C)$;
- (7) 分配律: $(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$;
- (8) 并、交的单调性: 如果 $A \subset B$, 则对任意的集合 C 有 $A \cup C \subset B \cup C, A \cap C \subset B \cap C$;
- (9) 差的分配律: $(A - B) \cap C = (A \cap C) - (B \cap C)$;
- (10) 如果 $A \subset B$, 则 $A - B = \emptyset$;
- (11) $(C - A) - B = C - (A \cup B)$;
- (12) 如果 $A \subset C, B \subset C$, 则 $A - B = A \cap C_C B$;
- (13) $A \cup B = (A \Delta B) \cup (A \cap B)$.

注: 如上的性质有些可以直接推广到任意多个集合的情形.

2.3.3 上、下限集与特征函数

1. 上、下限集

设 $A_1, A_2, \dots, A_n, \dots$ 是任意一列集合, 由属于上述集列中无限多个集合的元素全体组成的集合称为这一列集合的上限集, 记为 $\overline{\lim}_n A_n$. 而对于某个指标 n_0 , 由集列中所有满足 $n > n_0$ 的集合 A_n 的元素全体组成的集合称为集列的下限集, 记为 $\underline{\lim}_n A_n$. 显然有

$$\lim_n A_n \subset \overline{\lim_n A_n}.$$

事实上,由定义可以证明:

$$\overline{\lim_n A_n} = \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m, \quad \lim_n A_n = \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m.$$

如果集列 $\{A_n\}$ 的上限集与下限集相等,即 $\lim_n A_n = \overline{\lim_n A_n}$,则称集列 $\{A_n\}$ 收敛,称 $A = \lim_n A_n = \overline{\lim_n A_n}$ 是集列 $\{A_n\}$ 的极限(或极限集),记为 $A = \lim_{n \rightarrow \infty} A_n$.

如果集列 $\{A_n\}$ 是单调增加的,即满足 $A_n \subset A_{n+1} (n=1,2,\cdots)$,则 $\lim_{n \rightarrow \infty} A_n = \bigcup_{n=1}^{\infty} A_n$.

如果集列 $\{A_n\}$ 是单调减少的,即满足 $A_n \supset A_{n+1} (n=1,2,\cdots)$,则 $\lim_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} A_n$.

2. 集合的特征函数

设 X 是一个固定的非空集合, A 是 X 的一个子集. 作 X 上的函数

$$\chi_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases}$$

则称 $\chi_A(x)$ 为集合 A 的特征函数. 显然集合 A 完全由它的特征函数所确定,即当 $\chi_A(x) \equiv \chi_B(x)$ 时有 $A = B$.

集合与其特征函数之间有下列关系:

(1) $A = X$ 等价于 $\chi_A(x) \equiv 1$; $A = \emptyset$ 等价于 $\chi_A(x) \equiv 0$;

(2) $A \subset B$ 等价于 $\chi_A(x) \leq \chi_B(x)$;

(3) $A = B$ 等价于 $\chi_A(x) = \chi_B(x)$;

(4) 设 $A = \bigcup_{n=1}^{\infty} A_n, B = \bigcap_{n=1}^{\infty} A_n$, 则 $\chi_A(x) = \max_n \chi_{A_n}(x), \chi_B(x) = \min_n \chi_{A_n}(x)$.

2.4 合理分派与会成员问题^①

2.4.1 问题的提出

为了讨论重大问题,特别是像制定长远规划的问题,而召开小组讨论会正变得越来越普遍. 一般认为,由很多人参加的大会很难取得建设性的成果,甚至讨论结果容易被一位占支配地位的人控制或操纵. 因此,在公司召开董事会议时,应首先召集全体董事召开一些讨论有关事务的小组会议,这些较小规模的分组

^① 本案例选自 1997 年美国大学生数学建模竞赛 B 题,根据信息工程大学信息工程学院何王全,姜恺,宋献涛的获奖论文整理.

会议仍然有被某个占支配地位的人控制的危险. 为降低这种危险, 常用的办法是安排每个小组开几次会, 每次会有不同的人参加.

AnTostal 公司的一次会议的参加者为 29 位公司董事会成员, 其中 9 位是在职董事 (即公司的雇员). 会议要开一天, 每个小组上午开三段, 下午开四段, 每段会议开 45 分钟, 从上午 9:00 到下午 4:00 每整点开始开会, 中午 12:00 午餐. 上午的每段会议都有 6 个小组讨论会, 每个小组讨论会都由公司的一位资深高级职员来主持讨论. 这些资深高级职员都不是董事会的成员. 因此, 每位资深高级职员都要主持三个不同的小组讨论会. 这些资深高级职员不参加下午的讨论会, 而且下午的每段会议只有 4 个不同的小组讨论会.

公司董事长需要一份由公司董事参加的 7 段分组会议的每个小组的分配名单. 这份搭配名单要尽可能多地把董事均匀搭配. 理想的搭配应是任意两位董事同时参加一个小组讨论会的次数相同, 与此同时, 要使在不同时段的小组会中同在一起开过会的董事总数达到最小.

名单中的搭配还应满足下列两个准则:

(1) 在上午的讨论会上, 不允许一位董事参加由同一位资深高级职员主持的两次会议.

(2) 每个分组讨论会都应将在职董事均匀分配到各小组中.

给出一份 1~9 号在职董事、10~29 号董事、1~6 号公司资深高级职员的分组搭配名单, 说明该名单在多大程度上满足了前面提出的各种要求和规则, 因为有的董事可能在最后一分钟宣布不参加会议, 也可能不在名单上的董事将出席会议. 因此, 一个能使秘书在会前一小时接到参会与否的通知情况下, 来调整搭配分组的算法定会得到赏识. 如果算法还能用于不同水平的与会者与参加后面会议中的每一类与会者搭配的话, 那就更理想了.

2.4.2 模型的假设与符号说明

- (1) 各场会议间及各个小组之间是相对独立的;
- (2) 所有的资深高级职员和董事会成员都严格遵守派遣方案;
- (3) 若能够满足每位董事出席会议的次数都相等, 则模型被认为是最理想的;
- (4) 6 位资深高级职员之间无差异, 同样 9 位在职董事之间、20 位外部董事之间也无差异.

根据需要引入下列符号:

$O = \{o_i | i = 1, 2, \dots, 6\}$ 表示 6 位资深高级职员 $o_i (i = 1, 2, \dots, 6)$ 的集合;

$M = \{m_i | i = 1, 2, \dots, 29\}$ 表示所有董事会成员 $m_i (i = 1, 2, \dots, 29)$ 的集合;

$I(9) = \{m_i | i = 1, 2, \dots, 9\}$ 表示在职董事会成员 $m_i (i = 1, 2, \dots, 9)$ 的集合;

$E(20) = \{m_i | i = 10, 11, \dots, 29\} = M - I(9)$ 表示外部董事会成员 $m_i (i = 10, \dots, 29)$ 的集合;

G_n 表示在一次分组会议中第 $n (1 \leq n \leq 6)$ 组的与会成员的集合;

$G_n^{(k)}$ 表示分组会议的第 $n (1 \leq n \leq 6)$ 组经第 k 次分派 (每次分派一名成员) 后的会议成员集合;

a_{ij} 表示董事会第 i 位成员与第 j 位成员分在同一组的次数 $(1 \leq i, j \leq 29, i \neq j)$;

$w(k)$ 表示两位董事会成员分在同一组时所赋予的权重;

b_{ij} 表示资深高级职员 $o_i (i = 1, 2, \dots, 6)$ 与董事会成员 $m_j (j = 1, 2, \dots, 29)$ 在此之前是否同组的指标, 即当属于同一组时取值为 1, 否则取值为 0;

$R_i^{(k)} = |G_i^{(k)} \cap I(9)|$ 表示在 $G_i^{(k)}$ 中在职董事的数量;

$h_i/2$ 表示董事会两位在职成员在同一个讨论组中达到 i 次的对数;

t_i 表示第 i 组中在职董事数与外部董事数之比.

2.4.3 模型的分析

该问题需要我们对于安排不同的分组会议给出合理的分派方案, 以有效降低会议被某一个人员支配控制的风险, 从而确保讨论的充分性.

由上述假设, 我们可以容易知道, 每位资深高级职员或董事会成员被分派到任意会议讨论组的可能性是一样的, 同时, 他们中的任何人都不能根据其意愿选择会议讨论组. 根据问题要求, 我们总结出两个分派原则:

均衡分派原则: 尽可能使得各讨论组成员人数相等, 该原则将确保各讨论组人数大致相等.

分派比例原则: 各会议讨论组在职董事和外部董事的比例大致相等.

于是我们设想给出下面的两种分派方案:

方案 I:

如果公司总裁更希望使各讨论组的成员人数尽可能相等, 我们将采用这个方案, 即可满足他的要求. 显然, 这个问题的绝对理想的结果是不存在的, 但我们可以按如下方法使得结果尽可能地达到理想状态:

每场上午的分组会议共分为 6 组, 一个组由 4 位董事会成员组成, 其他 5 个组每组由 5 位董事会成员组成. 其中 3 个组每组有 2 位在职董事, 而另外 3 个组每组只有 1 位在职董事. 具体对于每一个组成员的分派都是随机的.

方案 II:

在实际中很多情况下, 不同类型与会者的比例是非常重要的, 基于这一点, 我们提出方案 II. 由均衡分派原则, 每组中在职董事和外部董事的比例应该近似相等, 即应大致为 $9/20 = 0.45$. 由所给数据和题目要求可知, 不存在恰好比例为

9/20 的分派方案,但我们可以从以下结果中挑选最接近的方案.

上午: $t_1 = 1:2, t_2 = 1:2, t_3 = 1:2, t_4 = 2:4, t_5 = 2:5, t_6 = 2:5$

或者 $t_1 = 1:2, t_2 = 1:2, t_3 = 1:3, t_4 = 2:4, t_5 = 2:4, t_6 = 2:5$;

下午: $t_1 = 2:4, t_2 = 2:4, t_3 = 2:5, t_4 = 3:7$.

2.4.4 模型的建立

模型 I: 第一种分派方案

第一步: 上午第一场会议的分派方案.

首先,随机地把 29 位董事会成员大致均匀地分成 6 组,其中一个组由 4 位成员组成,其他 5 个组每组由 5 位成员组成.然后,随机地将集合 O 中的 6 位资深高级职员分配到每一个组中,分别记为 $G_i (i = 1, 2, \dots, 6)$,从而完成了第一场会议的分派.

第二步: 上午第二场会议的分派方案.

首先,随机地任意取一个 $o_i \in O$,将其分到 G_i 中,即 $o_i \in G_i (i = 1, 2, \dots, 6)$,从而使每个组 G_i 都由一位资深高级职员 $o_i (i = 1, 2, \dots, 6)$ 作为主持.然后,按下面的方法分派董事会成员:

(1) 先为每一组 G_i 分派第一个董事会成员.对于任意一个 $m_{j_1} \in M (1 \leq j_1 \leq 29)$.若 $b_{j_1} = 0$,则令 $G_i^{(1)} = \{m_{j_1}\}$.否则,随机选择另一个 $m_{j_2} \in M (j_1 \neq j_2)$,直到 $b_{j_2} = 0$,并令 $G_i^{(1)} = \{m_{j_2}\} (i = 1, 2, \dots, 6)$.这样,就为每一个组分派了第一位董事会成员.

(2) 假设已为每一个组分派了 $k-1$ 位董事会成员,即 $G_i^{(k-1)} = \{\bar{m}_{j_1}, \bar{m}_{j_2}, \dots, \bar{m}_{j_{k-1}}\} (2 \leq k \leq 5)$ 已确定.要分派第 k 位董事会成员给 G_i ,即确定 $G_i^{(k)} = G_i^{(k-1)} \cup \{m_{j_k}\} (2 \leq k \leq 5)$.

任意随机选择一个 $m_{j_k} \in M - \bigcup_{i=1}^6 G_i^{(k-1)}$,计算 b_{j_k} 的值,根据 $b_{j_k} = 0$ 或 $b_{j_k} = 1$ 分别考虑.

1) 如果 $b_{j_k} = 0$,则考虑以下两种情况:

情况 1: 若 $m_{j_k} \in I(9)$,即为在职董事,并确定 $R_i^{(k-1)} (R_i^{(k-1)} \leq 2)$.

(a) 若 $R_i^{(k-1)} < 2$,则令集合 $G_i^{(k)} = \{\bar{m}_{j_1}, \bar{m}_{j_2}, \dots, \bar{m}_{j_{k-1}}, m_{j_k}\}$.

(b) 若 $R_i^{(k-1)} = 2$,则选择另外一位董事会成员 $m_{j_k} \in (M - \bigcup_{i=1}^6 G_i^{(k-1)}) \cap I(9)$,直到 $b_{j_k} = 0$ 且 $R_i^{(k-1)} < 2$ 为止,则令集合 $G_i^{(k)} = \{\bar{m}_{j_1}, \bar{m}_{j_2}, \dots, \bar{m}_{j_{k-1}}, m_{j_k}\}$.

情况 2: 如果 $m_{j_k} \notin I(9)$,即为外部董事,记

$$C = \{m_j \mid b_{j_k} = 0, m_j \in M - \bigcup_{i=1}^6 G_i^{(k-1)}, m_j \notin I(9)\}$$

为所有候选的外部董事的集合. 对于每个 $m_j \in C$, 计算 $q(m_j) = \sum_{m_i \in C} w(a_{ij})$, 在集合 C 中求出使得 $q(m_{j_k}) = \min_{m_j \in C} q(m_j)$ 的 m_{j_k} , 并且令集合

$$G_i^{(k)} = \{\bar{m}_{j_1}, \bar{m}_{j_2}, \dots, \bar{m}_{j_{k-1}}, m_{j_k}\}.$$

2) 如果 $b_{ij_k} = 1$, 则可以选择另一位董事会成员 $m_{j_k} \in M - \bigcup_{i=1}^6 G_i^{(k-1)}$, 直到 $b_{ij_k} = 0$, 同样使用上述方法, 也能确定集合 $G_i^{(k)}$. 最后, 我们可以确定出 $G_i = G_i^{(5)}$ ($i = 1, 2, \dots, 6$), 从而得到第二场会议的分派结果.

第三步: 类似上述方法分派上午第三场会议的分组结果.

第四步: 下午分组会议的分派方案安排.

(1) 随机任意选择 $m_{j_1}, m_{j_2}, m_{j_3}, m_{j_4} \in M$, 作为下午每一组的第一位成员.

(2) 重复上面的第二步和第三步, 注意在这里四个组的在职董事分配比例为 2:2:2:3. 这样, 我们就可以得到下午每场会议的分派方案.

模型 II: 第二种分派方案

根据上面分析, 我们采用下面的分派方案将会得到最佳的比例, 即有最理想的均衡性.

在上午每一场讨论会议的六个组中, 我们取在职董事与外部董事的比例分别为 (A): 1:2, 1:2, 1:2, 2:4, 2:5 和 2:5; 或者为 (B): 1:2, 1:2, 1:3, 2:4, 2:4 和 2:5; 在下午的会议四个组中, 两者的比例均为 2:4, 2:4, 2:5 和 3:7.

上午会议的分派方案:

首先考虑在职董事与外部董事的比例为 1:2, 1:2, 1:2, 2:4, 2:5 和 2:5 的情况. 将 6 名资深高级职员随机地分派到六个组中去.

(1) 按照如下的方法分派各讨论组的每一位成员:

从 9 位在职董事中随机选取 $m_{j_1} \in I(9)$, 如果 $b_{ij_1} = 0$, 则令 $G_i^{(1)} = \{m_{j_1}\}$; 否则, 另取 $m_{j_2} \in I(9)$ ($j_1 \neq j_2$), 并且 $b_{ij_2} = 0$, 则令 $G_i^{(1)} = \{m_{j_2}\}$.

(2) 将剩余的三位在职董事随机分配到任意三个组中, 就可以得到 $G_i^{(2)}$.

(3) 用模型 I 中第二步的方法进行分派, 我们给每个组分派两位外部董事.

(4) 将剩下的 $E(20)$ 中的 8 位成员, 依照 2:3:3 的比例分配到包含两位在职董事的三个组中去, 并保证 $R_i^{(4)} = 2$.

按照上面步骤, 我们就可以得到上午三场会议的分组方案.

对于在职董事与外部董事的比例为 1:2, 1:2, 1:3, 2:5, 2:4, 2:4 的情况, 上述方法仍适用. 只需注意在步骤 (4) 中, 将剩余 8 位外部董事中的 7 位依照 2:2:3 的比例分派到使 $R_i^{(4)} = 2$ 的三个组中去, 并分派最后一位到使 $R_i^{(4)} = 1$ 的组中去即可.

下午会议的分派方案:

在这里取在职董事与外部董事的比例分别为 2:4, 2:4, 2:5, 3:7。

(1) 随机地选取 4 位在职董事 $m_{j_1}, m_{j_2}, m_{j_3}, m_{j_4} \in I(9)$, 并令 $G_i^{(1)} = \{m_{j_i}\}$ 之后, 再选取 $m_{j'_1}, m_{j'_2}, m_{j'_3}, m_{j'_4} \in I(9) - \bigcup_{i=1}^4 G_i^{(1)}$, 使 $w(a_{j_i j'_i})$ 最小, 同时令 $G_i^{(2)} = \{m_{j_i}, m_{j'_i}\} (1 \leq j_i, j'_i \leq 9, \text{ 且 } j_i \neq j'_i, i = 1, 2, 3, 4)$. 最后, 将 $I(9)$ 中最后剩下的一位在职董事分派到能使 $w(a_{j_i j'_i})$ 最小的那一组中去.

(2) 同样应用模型 I 中的第二步方法, 将使 $R_i^{(3)} = 3$ 或者 1 的组做适当的调整使 $R_i^{(3)} = 2$.

(3) 将剩余 6 位外部董事用模型 I 中第二步的方法较均匀地分派到四个讨论组中去。

2.4.5 模型的求解结果

利用计算机模拟求解可以得到两个模型的求解结果,即分组讨论会的分派方案.

方案 I: $w(0)=0, w(1)=1, w(2)=3, w(3)=6, w(4)=40, w(5)=100$.
具体的分组方案如表 2-1 所示.

表 2-1 方案 I 的分组结果

第一场会议分组：

| | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|
| G_1 | O_1 | m_9 | m_{14} | m_{19} | m_{26} | m_{29} |
| G_2 | O_2 | m_1 | m_{20} | m_{21} | m_{27} | |
| G_3 | O_3 | m_3 | m_{10} | m_{22} | m_{24} | m_{28} |
| G_4 | O_4 | m_4 | m_7 | m_{11} | m_{15} | m_{16} |
| G_5 | O_5 | m_5 | m_6 | m_{12} | m_{17} | m_{23} |
| G_6 | O_6 | m_2 | m_8 | m_{13} | m_{18} | m_{25} |

第三场会议分组：

| | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|
| G_1 | O_1 | m_7 | m_8 | m_{17} | m_{21} | m_{24} |
| G_2 | O_2 | m_2 | m_{16} | m_{19} | m_{22} | m_{23} |
| G_3 | O_3 | m_5 | m_9 | m_{15} | m_{18} | |
| G_4 | O_4 | m_3 | m_{12} | m_{14} | m_{25} | m_{27} |
| G_5 | O_5 | m_1 | m_{11} | m_{13} | m_{26} | m_{28} |
| G_6 | O_6 | m_4 | m_6 | m_{10} | m_{20} | m_{29} |

第二场会议分组：

| | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|
| G_1 | O_1 | m_6 | m_{13} | m_{15} | m_{22} | m_{27} |
| G_2 | O_2 | m_4 | m_{12} | m_{18} | m_{24} | m_{26} |
| G_3 | O_3 | m_1 | m_7 | m_{23} | m_{25} | m_{29} |
| G_4 | O_4 | m_2 | m_9 | m_{17} | m_{20} | m_{28} |
| G_5 | O_5 | m_8 | m_{10} | m_{14} | m_{16} | |
| G_6 | O_6 | m_3 | m_5 | m_{11} | m_{19} | m_{21} |

第四场会议分组：

| | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|
| G_1 | m_2 | m_4 | m_{14} | m_{15} | m_{21} | m_{23} | m_{28} | |
| G_2 | m_5 | m_7 | m_{10} | m_{12} | m_{13} | m_{19} | m_{20} | m_{26} |
| G_3 | m_6 | m_8 | m_9 | m_{11} | m_{22} | m_{24} | m_{25} | |
| G_4 | m_1 | m_3 | m_{16} | m_{17} | m_{18} | m_{27} | m_{29} | |

第五场会议分组：

| | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|
| G_1 | m_3 | m_8 | m_{15} | m_{17} | m_{20} | m_{23} | m_{26} | |
| G_2 | m_1 | m_4 | m_9 | m_{12} | m_{13} | m_{21} | m_{22} | |
| G_3 | m_2 | m_5 | m_{10} | m_{11} | m_{24} | m_{27} | m_{29} | |
| G_4 | m_6 | m_7 | m_{14} | m_{16} | m_{18} | m_{19} | m_{25} | m_{28} |

第六场会议分组：

| | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|
| G_1 | m_1 | m_8 | m_{10} | m_{12} | m_{15} | m_{19} | m_{28} | |
| G_2 | m_3 | m_7 | m_{11} | m_{14} | m_{18} | m_{20} | m_{22} | m_{23} |
| G_3 | m_4 | m_5 | m_{13} | m_{17} | m_{24} | m_{25} | m_{27} | |
| G_4 | m_2 | m_6 | m_9 | m_{16} | m_{21} | m_{26} | m_{29} | |

方案Ⅱ： $w(0)=0, w(1)=1, w(2)=3, w(3)=6, w(4)=40, w(5)=100$.

(A) 上午分组比例：1:2, 1:2, 1:2, 2:4, 2:5, 2:5；下午分组比例：2:4, 2:4, 2:5, 3:7. 具体分组方案如表 2-2 所示.

(B) 上午分组比例：1:2, 1:2, 1:3, 2:4, 2:5, 2:5；下午分组比例：2:4, 2:4, 2:5, 3:7. 具体分组结果如表 2-3 所示.

表 2-2 方案Ⅱ的分组方式(A)的分组结果

第一场会议分组：

| | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|
| G_1 | O_1 | m_3 | m_6 | m_{12} | m_{17} | m_{22} | m_{25} | m_{29} |
| G_2 | O_2 | m_4 | m_5 | m_{13} | m_{18} | m_{23} | m_{26} | m_{28} |
| G_3 | O_3 | m_9 | m_{14} | m_{19} | | | | |
| G_4 | O_4 | m_7 | m_{15} | m_{20} | | | | |
| G_5 | O_5 | m_2 | m_{10} | m_{21} | | | | |
| G_6 | O_6 | m_1 | m_8 | m_{11} | m_{16} | m_{24} | m_{27} | |

第二场会议分组：

| | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|
| G_1 | O_1 | m_1 | m_7 | m_{13} | m_{19} | m_{21} | m_{24} | m_{28} |
| G_2 | O_2 | m_3 | m_9 | m_{15} | m_{16} | m_{22} | m_{27} | |
| G_3 | O_3 | m_6 | m_{10} | m_{23} | | | | |
| G_4 | O_4 | m_8 | m_{12} | m_{18} | | | | |
| G_5 | O_5 | m_4 | m_{11} | m_{17} | | | | |
| G_6 | O_6 | m_2 | m_5 | m_{14} | m_{20} | m_{25} | m_{26} | m_{29} |

第七场会议分组：

| | | | | | | | | |
|-------|-------|----------|----------|----------|----------|----------|----------|----------|
| G_1 | m_2 | m_{10} | m_{13} | m_{14} | m_{17} | m_{22} | m_{26} | |
| G_2 | m_3 | m_7 | m_{12} | m_{21} | m_{24} | m_{28} | m_{29} | |
| G_3 | m_4 | m_6 | m_8 | m_{11} | m_{18} | m_{19} | m_{23} | m_{27} |
| G_4 | m_1 | m_5 | m_9 | m_{15} | m_{16} | m_{20} | m_{25} | |

第三场会议分组：

| | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|
| G_1 | O_1 | m_2 | m_{15} | m_{18} | | | | |
| G_2 | O_2 | m_1 | m_{12} | m_{17} | | | | |
| G_3 | O_3 | m_3 | m_4 | m_{16} | m_{20} | m_{21} | m_{24} | |
| G_4 | O_4 | m_5 | m_6 | m_{11} | m_{19} | m_{25} | m_{27} | m_{28} |
| G_5 | O_5 | m_7 | m_8 | m_{14} | m_{22} | m_{23} | m_{26} | m_{29} |
| G_6 | O_6 | m_9 | m_{10} | m_{13} | | | | |

第四场会议分组：

| | | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|----------|
| G_1 | m_2 | m_5 | m_{12} | m_{16} | m_{27} | m_{29} | | | |
| G_2 | m_1 | m_4 | m_{14} | m_{15} | m_{23} | m_{25} | m_{28} | | |
| G_3 | m_8 | m_9 | m_{17} | m_{21} | m_{24} | m_{26} | | | |
| G_4 | m_3 | m_6 | m_7 | m_{10} | m_{11} | m_{13} | m_{18} | m_{19} | m_{20} |

第五场会议分组：

| | | | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|----------|----------|
| G_1 | m_6 | m_8 | m_{12} | m_{13} | m_{15} | m_{20} | m_{27} | | | |
| G_2 | m_1 | m_9 | m_{21} | m_{23} | m_{25} | m_{29} | | | | |
| G_3 | m_2 | m_4 | m_{16} | m_{19} | m_{22} | m_{26} | | | | |
| G_4 | m_3 | m_5 | m_7 | m_{10} | m_{11} | m_{14} | m_{17} | m_{18} | m_{24} | m_{28} |

第六场会议分组：

| | | | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|----------|----------|
| G_1 | m_2 | m_6 | m_{13} | m_{17} | m_{23} | m_{24} | | | | |
| G_2 | m_3 | m_8 | m_9 | m_{10} | m_{11} | m_{19} | m_{20} | m_{22} | m_{25} | m_{26} |
| G_3 | m_1 | m_4 | m_{14} | m_{15} | m_{18} | m_{29} | | | | |
| G_4 | m_5 | m_7 | m_{12} | m_{16} | m_{21} | m_{27} | m_{28} | | | |

表 2-3 方案 II 的分组方式 (B) 的分组结果

第一场会议分组：

| | | | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|--|--|
| G_1 | O_1 | m_4 | m_5 | m_{13} | m_{18} | m_{23} | m_{26} | m_{28} | | |
| G_2 | O_2 | m_9 | m_{14} | m_{19} | | | | | | |
| G_3 | O_3 | m_7 | m_{15} | m_{20} | | | | | | |
| G_4 | O_4 | m_1 | m_8 | m_{11} | m_{16} | m_{24} | m_{27} | | | |
| G_5 | O_5 | m_3 | m_6 | m_{12} | m_{17} | m_{22} | m_{25} | | | |
| G_6 | O_6 | m_2 | m_{10} | m_{21} | m_{29} | | | | | |

第二场会议分组：

| | | | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|--|--|
| G_1 | O_1 | m_3 | m_9 | m_{15} | m_{16} | m_{22} | m_{27} | | | |
| G_2 | O_2 | m_6 | m_{10} | m_{23} | | | | | | |
| G_3 | O_3 | m_8 | m_{12} | m_{18} | | | | | | |
| G_4 | O_4 | m_2 | m_5 | m_{14} | m_{20} | m_{25} | m_{29} | | | |
| G_5 | O_5 | m_1 | m_7 | m_{13} | m_{19} | m_{21} | m_{24} | m_{28} | | |
| G_6 | O_6 | m_4 | m_{11} | m_{17} | m_{26} | | | | | |

第三场会议分组：

| | | | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|--|--|
| G_1 | O_1 | m_1 | m_{12} | m_{17} | | | | | | |
| G_2 | O_2 | m_3 | m_4 | m_{16} | m_{20} | m_{21} | m_{23} | m_{24} | | |
| G_3 | O_3 | m_5 | m_6 | m_{11} | m_{19} | m_{25} | m_{27} | | | |
| G_4 | O_4 | m_9 | m_{10} | m_{13} | m_{28} | | | | | |
| G_5 | O_5 | m_2 | m_{15} | m_{18} | | | | | | |
| G_6 | O_6 | m_7 | m_8 | m_{14} | m_{22} | m_{26} | m_{29} | | | |

第七场会议分组：

| | | | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|----------|----------|
| G_1 | m_4 | m_5 | m_{12} | m_{22} | m_{24} | m_{25} | | | | |
| G_2 | m_1 | m_7 | m_{10} | m_{20} | m_{23} | m_{26} | m_{27} | | | |
| G_3 | m_2 | m_3 | m_8 | m_{13} | m_{15} | m_{17} | m_{18} | m_{19} | m_{21} | m_{29} |
| G_4 | m_6 | m_9 | m_{11} | m_{14} | m_{16} | m_{28} | | | | |

第四场会议分组：

| | | | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|----------|----------|
| G_1 | m_1 | m_9 | m_{21} | m_{23} | m_{25} | m_{29} | | | | |
| G_2 | m_2 | m_3 | m_7 | m_{10} | m_{11} | m_{13} | m_{14} | m_{17} | m_{24} | m_{28} |
| G_3 | m_6 | m_8 | m_{12} | m_{15} | m_{18} | m_{20} | m_{27} | | | |
| G_4 | m_4 | m_5 | m_{16} | m_{19} | m_{22} | m_{26} | | | | |

第五场会议分组：

| | | | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|----------|----------|
| G_1 | m_2 | m_3 | m_8 | m_{11} | m_{13} | m_{19} | m_{20} | m_{22} | m_{25} | m_{26} |
| G_2 | m_6 | m_9 | m_{10} | m_{17} | m_{23} | m_{24} | | | | |
| G_3 | m_5 | m_7 | m_{12} | m_{16} | m_{21} | m_{27} | m_{28} | | | |
| G_4 | m_1 | m_4 | m_{14} | m_{15} | m_{18} | m_{29} | | | | |

第六场会议分组：

| | | | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|----------|----------|
| G_1 | m_2 | m_8 | m_{10} | m_{20} | m_{23} | m_{26} | m_{27} | | | |
| G_2 | m_6 | m_9 | m_{17} | m_{21} | m_{28} | m_{29} | | | | |
| G_3 | m_1 | m_3 | m_7 | m_{13} | m_{14} | m_{15} | m_{16} | m_{18} | m_{19} | m_{24} |
| G_4 | m_4 | m_5 | m_{11} | m_{12} | m_{22} | m_{25} | | | | |

第七场会议分组：

| | | | | | | | | | | |
|-------|-------|-------|----------|----------|----------|----------|----------|----------|----------|----------|
| G_1 | m_5 | m_6 | m_8 | m_{10} | m_{11} | m_{12} | m_{13} | m_{16} | m_{21} | m_{25} |
| G_2 | m_1 | m_3 | m_{14} | m_{15} | m_{18} | m_{24} | m_{26} | | | |
| G_3 | m_2 | m_9 | m_{17} | m_{23} | m_{27} | m_{28} | | | | |
| G_4 | m_4 | m_7 | m_{19} | m_{20} | m_{22} | m_{29} | | | | |

2.4.6 模型的检验与分析

方案 I:

(1) 由计算机模拟检验的结果显示: $w(k)$ ($k=0,1,\dots,5$) 的变化对 $h_i/2$ 有

一定的影响,但对期望 $E = \frac{\sum_{i=0}^7 ih_i}{29 \times 28}$ 和方差 $D = \frac{\sum_{i=0}^7 (i-E)^2 \times h_i}{28 \times 29}$ 几乎没有影响.

例如:

如果 $w(0)=0, w(1)=1, w(2)=3, w(3)=6, w(4)=40, w(5)=80$, 则计算结果为 $h_0=90, h_1=384, h_2=334, h_3=4, h_4=h_5=h_6=h_7=0, E \approx 1.310, D \approx 0.445$.

如果 $w(0)=0, w(1)=1, w(2)=6, w(3)=20, w(4)=30, w(5)=50$, 则计算结果为 $h_0=79, h_1=434, h_2=291, h_3=15, h_4=h_5=h_6=h_7=0, E \approx 1.306, D \approx 0.442$.

上面的结果表明方案 I 的稳定性令人较满意.

(2) 通过大量的计算机模拟可以发现: h_3, h_4, h_5, h_6, h_7 都比较接近 0, h_0 也较小, 而 h_1, h_2 稍大, 这也表明按现在题目的要求方案是理想的. 而当与会者的人数较多, 但分组数量较小时, 该方案就不是十分理想了.

方案 II:

(1) 通过 1 000 次的计算机仿真, 我们发现 $w(k)$ 的变化对 $h_i/2$ 几乎没有影响.

如果 $w(0)=0, w(1)=1, w(2)=3, w(3)=6, w(4)=40, w(5)=80$, 则计算结果为 $h_0=98, h_1=326, h_2=334, h_3=48, h_4=h_5=h_6=h_7=0, E \approx 1.40, D \approx 0.75$.

如果 $w(0)=0, w(1)=1, w(2)=6, w(3)=30, w(4)=50, w(5)=100$, 则计算结果为 $h_0=110, h_1=320, h_2=332, h_3=50, h_4=h_5=h_6=h_7=0, E \approx 1.41, D \approx 0.74$. 这足以说明这个方案是稳定的, 也是理想的.

(2) 方案 II 在社会活动中有广泛的应用, 适当的修改方案中的派遣方式, 就可以使之适用于各种不同类型的会议分组问题. 如果与会者数目较大而分组数量较小, 该方案较方案 I 理想.

2.4.7 模型的进一步讨论

如果出现了某些人临时不能到会, 或者某些人事先没有预约来参加会议的情况, 我们可以采用下面的方法来调整会议的分组.

(1) 一个理事会成员 $m_{i_0} \in G_i$ 因故不能出席分组会议.

方案 I:

若 $m_{i_0} \in I(9)$ 且 $(G_i - \{m_{i_0}\}) \cap I(9) = \emptyset$, 则从 G_j ($|G_j \cap I(9)| = 2$) 中随机

选取 $m_{j_0} \in G_j \cap I(9)$, 将其安排到 G_i 中. 否则, 即 $|(G_i - \{m_{j_0}\}) \cap I(9)| = 1$, 则无需调整.

若 $m_{j_0} \in E(20)$ 且 $|(G_i - \{m_{j_0}\}) \cap I(9)| = 2$, 则从 G_j ($|G_j \cap I(9)| = 1$) 中随机选取 $m_{j_0} \in G_j \cap E(20)$, 将其安排到 G_i 中. 否则就无需调整.

方案 II:

若 $m_{j_0} \in I(9)$, 且 $G_i - \{m_{j_0}\}$ 使 $0 < t_i < \frac{9}{20}$, 则任意选取一个 $m_{i'} \in G_i \cap E(20)$, 将其安排到使 $t_j > \frac{9}{20}$ 的组 G_j 中. 或者任取 $m_{j_0} \in G_j \cap I(9)$ 安排到 G_i 中.

若 $m_{j_0} \in E(20)$, 则可作类似的调整.

(2) 出现了一个事先没有预约的成员(用 m_r 表示)来参加会议.

方案 I:

若 m_r 是在职董事, 则将其安排到 G_i 中使得 $|G_i \cap I(9)| = 1$. 同时, 我们将从 $G_j \cap E(20)$ ($j \neq i$) 中随机选取 m_{j_0} , 并将其安排到 G_i 中.

若 m_r 不是在职董事, 则将其安排到使 $|G_j \cap E(20)| = 3$ 的组 G_j 中, 或者将其安排到使 $|G_i \cap I(9)| = 2$ 的组 G_i 中.

方案 II:

若 m_r 是在职董事, 将其安排到使在职董事与外部董事之比 t_j 最小的组 G_j 中, 否则, 将其安排到使在职董事与外部董事之比 t_j 最大的组 G_j 中.

当有两个或更多个的这种情况时, 处理的方法依次类推.

2.5 参考案例与参考文献

1. 参考案例

- (1) 机 - 电系统的类比问题——文献[2]:18—20
- (2) 航船阻力问题——文献[3]:105—107
- (3) 核军备竞赛问题——文献[2]:37—38
- (4) 传染病的随机传染问题——文献[1]:48—52
- (5) 赛程安排问题——文献[3]:146—148

2. 参考文献

- [1] 姜启源. 数学模型. 第二版. 北京: 高等教育出版社, 1993
- [2] 陈义华. 数学模型. 重庆: 重庆大学出版社, 1995
- [3] 徐全智, 杨晋浩. 数学建模. 北京: 高等教育出版社, 2003
- [4] 叶其孝. 大学生数学建模竞赛辅导教材(四). 长沙: 湖南教育出版社, 2001

第三章 微分方程方法

3.1 微分方程的一般理论

微分方程是研究函数变化规律的有力工具,有着广泛的实际应用.针对所研究的对象建立微分方程模型是解决问题的第一步,实际中只有求出微分方程的解才能对所研究的问题进行解释说明.一般说来,求微分方程的解析解是困难的,大多数的微分方程需要用数值方法来求解,因此首先需要研究微分方程的解的存在惟一性和稳定性问题.

3.1.1 微分方程的一般形式

一阶微分方程

$$\begin{cases} \frac{dx}{dt} = f(t, x), \\ x(t_0) = x_0, \end{cases} \quad (3.1)$$

其中 $f(t, x)$ 是 t 和 x 的已知函数, $x(t_0) = x_0$ 为初始条件,又称定解条件.

一阶微分方程组

$$\begin{cases} \frac{dx_i}{dt} = f_i(t, x_1, x_2, \dots, x_n) \quad (i = 1, 2, \dots, n), \\ x_i(t_0) = x_i^{(0)} \quad (i = 1, 2, \dots, n) \end{cases} \quad (3.2)$$

又称为一阶正规方程组. 如果引入向量

$$\mathbf{x} = (x_1, x_2, \dots, x_n)^T, \mathbf{x}_0 = (x_1^{(0)}, x_2^{(0)}, \dots, x_n^{(0)})^T,$$

$$\mathbf{f} = (f_1, f_2, \dots, f_n)^T, \frac{d\mathbf{x}}{dt} = \left(\frac{dx_1}{dt}, \frac{dx_2}{dt}, \dots, \frac{dx_n}{dt} \right)^T,$$

则方程组(3.2)可以写为简单的形式

$$\begin{cases} \frac{d\mathbf{x}}{dt} = \mathbf{f}(t, \mathbf{x}), \\ \mathbf{x}(t_0) = \mathbf{x}_0, \end{cases} \quad (3.3)$$

即与方程(3.1)的形式相同,当 $n=1$ 时为方程(3.1).

对于任一高阶的微分方程

$$\frac{d^n x}{dt^n} = f\left(t; x, \frac{dx}{dt}, \dots, \frac{d^{n-1}x}{dt^{n-1}}\right),$$

如果记 $\frac{d^i x}{dt^i} = y_i (i=0, 1, 2, \dots, n)$, 则方程为 $\frac{dy_{n-1}}{dt} = f(t; y_0, y_1, \dots, y_{n-1})$, 即可化为一阶方程组的形式. 因此, 下面主要对正规方程组 (3.3) 进行讨论.

3.1.2 微分方程解的存在惟一性

正规方程组 (3.3) 的解在什么条件下存在, 且惟一呢? 有下面的定理.

定理 3.1 (Cauchy - Peano) 如果函数 $f(t, x)$ 在 $R: |t - t_0| \leq a, |x - x_0| \leq b$ 上连续, 则方程组 (3.3) 在 $|t - t_0| \leq h$ 上有解 $x = \phi(t)$ 满足初值条件 $x_0 = \phi(t_0)$, 此处 $h = \min\left(a, \frac{b}{M}\right)$, $M = \max_{(t,x) \in R} |f(t, x)|$.

定理 3.2 如果函数 $f(t, x)$ 在 $R: |t - t_0| \leq a, |x - x_0| \leq b$ 上连续, 且满足利普希茨 (Lipschitz) 条件 (即存在正常数 L 使得 $|f(t, x^{(1)}) - f(t, x^{(2)})| \leq L|x^{(1)} - x^{(2)}|$, 其中 $(t, x^{(1)}), (t, x^{(2)}) \in R$), 则方程组 (3.3) 满足初值条件 $x_0 = \phi(t_0)$ 的解是惟一的.

定理证明详略^[8].

3.1.3 微分方程的稳定性问题

在实际问题中, 微分方程所描述的是物质系统的运动规律, 在用微分方程来研究这个物理过程中, 人们只能考虑影响该过程的主要因素, 而不得不忽略一些认为次要的因素, 这种次要的因素通常称为干扰因素. 这些干扰因素在实际中可以瞬时地起作用, 也可持续地起作用. 从数学上来看, 前者会引起初值条件的变化, 而后者则会引起微分方程本身的变化. 在实际问题中, 干扰因素是客观存在的, 由此可见, 对于它的影响程度的研究是必要的, 即初值条件或微分方程的微小变化是否也只引起对应解的微小变化? 这就是微分方程的稳定性问题. 这里仍以方程组 (3.3) 为例讨论.

1. 有限区间的稳定性

如果 $f(t, x)$ 在某个有限的区域 $G \subset \mathbb{R}^{n+1}$ 内连续, 且对 x 满足利普希茨条件, $x = \psi(t) (a \leq t \leq b)$ 是方程组 (3.3) 的一个特解, 则当 x_0 充分接近于 $\psi(t_0) (a \leq t_0 \leq b)$ 时, 方程组 (3.3) 在 $a \leq t \leq b$ 上满足初值条件 $x_0 = x(t_0)$ 的解 $x = \phi(t, t_0, x_0)$ 有

$$\lim_{x_0 \rightarrow \psi(t_0)} \phi(t, t_0, x_0) = \psi(t) \quad (a \leq t \leq b),$$

即对任意给定的 $\varepsilon > 0$, 总存在相应的 $\delta(\varepsilon) > 0$, 当 $|x_0 - \psi(t_0)| < \delta(\varepsilon)$ 时, 对一切 $a \leq t \leq b$ 有

$$|\phi(t, t_0, x_0) - \psi(t)| < \varepsilon,$$

此时称方程组(3.3)的解 $x = \psi(t)$ 在有限区间 $a \leq t \leq b$ 上是稳定的.

2. 无限区间的稳定性

如果 $x = \psi(t)$ ($t \geq t_0$) 是方程组(3.3)的一个特解, $x = \phi(t, t_0, x_0)$ ($t \geq t_0$) 是方程组(3.3)满足初值条件 $x_0 = x(t_0)$ 的解. 对任意给定的 $\varepsilon > 0$, 总存在相应的 $\delta(\varepsilon) > 0$, 当 $|x_0 - \psi(t_0)| < \delta(\varepsilon)$ 时, 对一切 $t \geq t_0$ 有

$$|\phi(t, t_0, x_0) - \psi(t)| < \varepsilon,$$

则称方程组(3.3)的解 $x = \psi(t)$ 在无限区间 $t \geq t_0$ 上是稳定的, 即无限区间上的稳定.

3. 渐近稳定性

如果方程组(3.3)解 $x = \psi(t)$ 在无限区间 $t \geq t_0$ 上是稳定的, 且存在 $\delta_0 > 0$, 当 $|x_0 - \psi(t_0)| < \delta_0$ 时, 有

$$\lim_{t \rightarrow \infty} (\phi(t, t_0, x_0) - \psi(t)) = 0,$$

则称 $x = \psi(t)$ 是渐近稳定的, 或称局部渐近稳定性.

如果上述 $\delta_0 = \infty$ (或给定的一个有限常数), 则相应的渐近稳定性称为全局渐近稳定性 (或大范围渐近稳定性).

4. 经常扰动下的稳定性

对于方程组(3.3), 考虑相应的方程组

$$\frac{dx}{dt} = f(t, x) + R(t, x), \quad (3.4)$$

这里的 $R(t, x)$ 称为扰动函数.

如果对任意给定的 $\varepsilon > 0$, 总存在 $\delta(\varepsilon) > 0$ 和 $\eta(\varepsilon) > 0$, 使得当 $|x_0 - \psi(t_0)| < \delta(\varepsilon)$ 时有

$$|R(t, x)| < \eta(\varepsilon),$$

则方程组(3.4)有满足初值条件 $x_0 = x(t_0)$ 的解 $x = \phi(t, t_0, x_0)$ ($t \geq t_0$). 且当 $t \geq t_0$ 时有

$$|\phi(t, t_0, x_0) - \psi(t)| < \varepsilon$$

就说方程组(3.3)的特解 $x = \psi(t)$ 在经常扰动下是稳定的.

5. 研究稳定性的方法

实际中, 要研究方程组(3.3)的解 $x = \psi(t)$ 的稳定性问题. 可以转化为研究方程的零解 (平凡解) 的稳定性问题. 事实上:

对于方程组(3.3)的任一特解 $x = \psi(t)$, 只要令 $y = x - \psi(t)$, 则

$$\begin{aligned} \frac{dy}{dt} &= \frac{dx}{dt} - \frac{d\psi(t)}{dt} = f(t, x) - f(t, \psi(t)) \\ &= f(t, y + \psi(t)) - f(t, \psi(t)) = g(t, y). \end{aligned}$$

显然有 $g(t, 0) \equiv 0$. 故方程组 (3.3) 变为

$$\frac{dy}{dt} = g(t, y). \quad (3.5)$$

于是可知方程组 (3.3) 的解 $x = \psi(t)$ 对应于方程组 (3.5) 为 $y = 0$ (平凡解). 因此, 要研究方程组 (3.3) 的 $x = \psi(t)$ 的稳定性问题可转化为研究方程组 (3.5) 的平凡解 $y = 0$ 的稳定性问题.

如果微分方程组的所有解都能简单地求出来, 一个特解的稳定性问题并不难解决. 然而, 实际中这种情况太少了. 因此, 一般性的稳定性问题的研究是复杂的, 通常的情况下都是针对具体问题做相应的研究.

3.2 微分方程的平衡点及稳定性

3.2.1 微分方程的平衡点

设有微分方程组 (3.3), 对于 $x = (x_1, x_2, \dots, x_n)^T \in \mathbf{R}^n, t \in [a, b], f(t, x)$ 在某个区域内连续, 且满足解的存在惟一性条件. 如果存在某个常数 $x_0 \in \mathbf{R}^n$, 使得 $f(t, x_0) = 0$, 则称点 x_0 为方程组 (3.3) 的平衡点 (或奇点), 且称 $x = x_0$ 为方程组的平凡解 (或奇解).

如果对所有可能初值条件, 方程组 (3.3) 的解 $x = \psi(t)$ 都满足

$$\lim_{t \rightarrow \infty} \psi(t) = x_0,$$

则称平衡点 x_0 是稳定的 (渐近稳定); 否则是不稳定的.

实际中, 判断平衡点的稳定性有两种方法: 间接方法和直接方法^[3].

间接方法: 首先求出方程的解 $x = \psi(t)$, 然后利用定义 $\lim_{t \rightarrow \infty} \psi(t) = x_0$ 来判断.

直接方法: 不用求方程的解直接地来研究其稳定性.

3.2.2 一阶方程的平衡点及稳定性

设有微分方程 $\frac{dx}{dt} = f(x)$, 其相应的平衡点为代数方程 $f(x) = 0$ 的实根 $x = x_0$. 其稳定性可以用间接方法判断, 下面说明直接方法.

首先, 将函数 $f(x)$ 在 x_0 点作一阶泰勒 (Taylor) 展开, 即方程可以近似地表示为

$$\frac{dx}{dt} = f'(x_0)(x - x_0).$$

显然, x_0 也是该方程的一个平衡点, 其稳定性主要取决于 $f'(x_0)$ 符号, 即有下面结论:

若 $f'(x_0) < 0$, 则平衡点 x_0 是稳定的; 若 $f'(x_0) > 0$, 则平衡点是不稳定的.

3.2.3 平面方程的平衡点及稳定性

设平面方程组的一般形式为

$$\begin{cases} \frac{dx_1}{dt} = f(x_1, x_2), \\ \frac{dx_2}{dt} = g(x_1, x_2), \end{cases} \quad (3.6)$$

则称代数方程组

$$\begin{cases} f(x_1, x_2) = 0, \\ g(x_1, x_2) = 0 \end{cases}$$

的实根 $x_1 = x_1^{(0)}, x_2 = x_2^{(0)}$ 为平面方程组 (3.6) 的平衡点, 记为 $P_0(x_1^{(0)}, x_2^{(0)})$. 如果对所有可能的初值条件方程的解为 $x_1(t), x_2(t)$ 满足

$$\lim_{t \rightarrow \infty} x_1(t) = x_1^{(0)}, \lim_{t \rightarrow \infty} x_2(t) = x_2^{(0)},$$

则称平衡点 $P_0(x_1^{(0)}, x_2^{(0)})$ 是稳定的; 否则是不稳定的. 也可以用直接方法讨论.

将方程组 (3.6) 的右边的函数作一阶泰勒展开, 即可表示为近似的线性方程组

$$\begin{cases} \frac{dx_1}{dt} = f_{x_1}(x_1^{(0)}, x_2^{(0)})(x_1 - x_1^{(0)}) + f_{x_2}(x_1^{(0)}, x_2^{(0)})(x_2 - x_2^{(0)}), \\ \frac{dx_2}{dt} = g_{x_1}(x_1^{(0)}, x_2^{(0)})(x_1 - x_1^{(0)}) + g_{x_2}(x_1^{(0)}, x_2^{(0)})(x_2 - x_2^{(0)}), \end{cases} \quad (3.7)$$

记系数矩阵为 $A = \begin{bmatrix} f_{x_1} & f_{x_2} \\ g_{x_1} & g_{x_2} \end{bmatrix}_{P_0}$, 且假设其行列式 $|A| \neq 0$, 则方程组 (3.7) 的特征

方程为

$$|A - \lambda I| = 0, \text{ 即 } \lambda^2 + p\lambda + q = 0,$$

其中 $p = -(f_{x_1} + g_{x_2})|_{P_0}, q = |A|, \lambda$ 为特征根. 不妨设特征根分别为 λ_1, λ_2 , 即

$$\lambda_1, \lambda_2 = \frac{1}{2}(-p \pm \sqrt{p^2 - 4q}).$$

根据特征根 λ_1, λ_2 和系数 p, q 的取值情况可以确定平衡点 $P_0(x_1^{(0)}, x_2^{(0)})$ 的稳定性.

事实上, 当 $p > 0, q > 0$ 时平衡点是稳定的; 当 $p < 0$ 或 $q < 0$ 时平衡点是不稳定的.

对于一般微分方程的平衡点和稳定性问题可以类似地讨论.

3.3 战争的预测与评估问题

3.3.1 问题的提出

目前,在超级大国的全球战略的影响下,世界并不太平,国与国之间和地区之间的种族歧视、民族矛盾、利益冲突、历史遗留问题等原因造成的局部战争和地区性武装冲突时有发生,有的长期处于敌对状态,从而导致了地区性的紧张局势和潜在的战争威胁.在这种情况下,必然会导致敌对双方的军备竞赛,在一定的条件下就会爆发战争.随着高科技的发展,尤其是信息技术的发展,军事装备现已成为决定战争胜负重要因素.在这里我们所说的军事装备是指军事实力的总和,主要包括武器装备、电子信息装备、军事兵力、军事费用等.

现代条件下的战争,一般都是多兵种的协同作战,所谓的多兵种就是综合使用陆、海、空、导弹、空降等兵力和相应的武器装备去完成不同的战争任务.由于每一兵种和相应的武器装备都有各自的优势和相应的适合攻击的目标.因此,现代战争的结局在很大程度上取决于是否能够广泛合理地利用诸兵种的合成部队协同作战,在战争中争取保持一定优势,尤其是在“制空权”和“制海权”的优势,这是现代战争的一大特点.

另一方面,现代战争往往是根据不同兵种的特点,可以在不同的区域参加战斗,即一场战争可以在不同几个区域同时展开,都对战争的结果产生一定的影响.

现在要求建立数学模型讨论的以下问题:

- (1) 分析研究引起军备竞赛的因素,并就诸多因素之间的相互关系进行讨论;
- (2) 在多兵种的作战条件下,对作战双方的战势进行评估分析.

3.3.2 模型的假设

- (1) 敌对双方为甲方和乙方,时刻 t 的军备综合实力分别为 $x(t)$ 和 $y(t)$;
- (2) 双方的军备综合实力是随着时间连续平稳变化的,即 $x(t)$ 和 $y(t)$ 是时间 t 的连续可微函数;
- (3) 不考虑第三方的军备实力对甲乙双方的影响.

3.3.3 模型的建立与求解

问题(1):

根据实际情况,一般认为促使和制约敌对双方的军备竞赛的因素主要有双

方各自的固有增长因素、双方敌对的程度和现有的军备实力等因素。

首先,由于各自的历史地位、地理环境和领土争端等原因,双方都有一个固有的增加军备的需求,即各自的固有军备增长率,分别记为常数 α 和 β 。

其次,双方的军备增长与双方的敌对程度有关,即随着敌对情绪的增长而增加。如果一方的军备增加了,则另一方也必然要增加自己的军备,以至于要赶上或超过对方。即甲方的军备实力的增长与乙方的军备实力成正比,反之亦然。其比例系数分别记为 a 和 b ,即表示受对方现有军备实力的刺激程度的度量。

再次,各方军备的增长与现有军备实力有关,由于经济实力的制约作用,军备实力越大,受经济制约的程度就越大,即军备增长率减少的程度与现有的军备实力成正比,其比例系数分别记为 c 和 d ,即表示双方受各自经济制约程度的度量。

于是,可以得到甲乙双方的军备实力的增长率变化情况,即军备竞赛的数学模型为

$$\begin{cases} \frac{dx}{dt} = -cx + ay + \alpha, \\ \frac{dy}{dt} = bx - dy + \beta. \end{cases} \quad (3.8)$$

为了要研究军备竞赛的结局,我们来求(3.8)式的平衡点,即令

$$\begin{cases} -cx + ay + \alpha = 0, \\ bx - dy + \beta = 0, \end{cases}$$

可以解得平衡点为

$$x^* = \frac{d\alpha + a\beta}{cd - ab}, \quad y^* = \frac{b\alpha + c\beta}{cd - ab} \quad (cd \neq ab).$$

根据平衡点的稳定性理论可知:当 $cd > ab$ 时,平衡点 (x^*, y^*) 是稳定的,否则是不稳定的。这就意味着在足够长的时间以后,双方的军备实力会分别达到一个稳定的极限值。

当 $cd > ab$ 时,方程(3.8)的平衡点 (x^*, y^*) 稳定,即说明当双方制约发展军备的程度大于刺激对方发展军备的程度时,军备竞赛的最终结果是可以达到平衡的。相反的,当 $cd < ab$ 时,方程组(3.8)的平衡点 (x^*, y^*) 不稳定,即说明当双方制约发展军备的程度小于刺激对方发展军备的程度时,双方的军备竞赛会一直无限地进行下去,最终会导致战争。

当 $\alpha = \beta = 0$, 且 $cd > ab$ 时,方程(3.8)的平衡点 $(x^*, y^*) = (0, 0)$ 是稳定的。即说明甲乙双方没有利害冲突和争端,在和平共处的情况下,都没有发展军备的欲望。

当 $\alpha \neq 0, \beta \neq 0$, 且 $cd > ab$ 时, 即表明了双方军备竞赛的存在性, 即便是因为某种外界因素的影响, 迫使双方在某个时候有 $x(t) = 0$ 和 $y(t) = 0$ (被迫裁军), 但由于 $\frac{dx}{dt} = \alpha > 0$ 和 $\frac{dy}{dt} = \beta > 0$, 则双方的军备竞赛客观存在, 最终双方的军备实力还会强大起来. 此时, 平衡点是稳定的, 所以最终还是会达到平衡.

如果有某种原因, 迫使某一方单方面裁军, 譬如对甲方来说, 即使在某个时候有 $x(t) = 0$, 但由于 $\frac{dx}{dt} = ay + \alpha$, 即由于乙方军备的存在, 对甲方有一定的刺激作用, 以及甲方有固有的军备增长需求, 则甲方的军备很快还会发展起来, 这说明单方面的裁军是不会长久的.

问题(2):

在由多兵种的协同作战的情况下, 一般认为甲乙双方的每一兵种(或一类武器装备, 或作战单位等)都有自己确定的作战目标. 为此, 我们假设双方的目标分配都已确定, 为了各自的目标去争取最好的作战效果.

不妨设甲方和乙方分别有 m 个和 n 个兵种(或作战单位、武器装备类), 其数量分别用向量 x 和 y 表示, 即

$$x(t) = (x_1(t), x_2(t), \dots, x_m(t))^T, \quad y(t) = (y_1(t), y_2(t), \dots, y_n(t))^T,$$

其中 $x_i(t)$ 表示 t 时刻甲方第 i 个兵种的数量, $y_j(t)$ 表示 t 时刻乙方第 j 个兵种的数量.

在战斗过程中, 双方的任何一个兵种对对方任何一个兵种都会构成一定的威胁, 也会被对方造成一定损失. 用 a_{ji} 表示 x_i 对 y_j 的损耗系数, 即 x_i 对 y_j 的战斗力; b_{ij} 表示 y_j 对 x_i 的损耗系数, 即 y_j 对 x_i 的战斗力. 通常情况下都有 $a_{ji} > 0, b_{ij} > 0$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$). 实际中, 甲方的任何一个兵种(或作战单位)都可攻击乙方的任何一个兵种(或作战单位), 反之亦然. 用 α_{ji} 表示 x_i 用于攻击 y_j 的比例(或甲方的第 i 个兵种用于攻击乙方的第 j 个兵种的概率); β_{ij} 表示 y_j 用于攻击 x_i 的比例(或乙方的第 j 个兵种用于攻击甲方的第 i 个兵种的概率). 于是可以得到多兵种作战的数学模型为

$$\begin{cases} \frac{dx_i}{dt} = - \sum_{j=1}^n b_{ij} \beta_{ij} y_j, & i = 1, 2, \dots, m, \\ \frac{dy_j}{dt} = - \sum_{i=1}^m a_{ji} \alpha_{ji} x_i, & j = 1, 2, \dots, n. \end{cases} \quad (3.9)$$

这就是著名的兰彻斯特(Lanchester)多兵种作战模型.

为了方便, 引入矩阵记号

$$A = (a_{ji})_{n \times m}, \quad B = (b_{ij})_{m \times n}, \quad P = (\alpha_{ji})_{n \times m}, \quad Q = (\beta_{ij})_{m \times n},$$

则模型(3.9)可以表示为矩阵形式

$$\begin{cases} \frac{dx}{dt} = -(B * Q) \cdot y, \\ \frac{dy}{dt} = -(A * P) \cdot x, \end{cases}$$

其中 $A * P = (a_{ji}\alpha_{ji})_{n \times m}$ 与 $B * Q = (b_{ij}\beta_{ij})_{m \times n}$ 都表示两个同阶矩阵的对应元素相乘以后的矩阵.

在多兵种协同作战的情况下, 战斗中甲乙双方的任何一个兵种(或作战单位)都可能受到不同程度的损失, 为了争取作战的优势, 每一个兵种都有可能在不断地补充一些兵力(或武器装备). 分别用 r_i 和 q_j 表示甲方的第 i 个兵种的兵力补充系数和乙方的第 j 个兵种的兵力补充系数. 则相应模型变为

$$\begin{cases} \frac{dx_i}{dt} = - \sum_{j=1}^n b_{ij}\beta_{ij}y_j + r_i, & i = 1, 2, \dots, m, \\ \frac{dy_j}{dt} = - \sum_{i=1}^m a_{ji}\alpha_{ji}x_i + q_j, & j = 1, 2, \dots, n. \end{cases}$$

该模型充分考虑了现代战争的多兵种的协同作战的特点, 但没有更多地考虑 C^3I 或 C^4I (Communication, Command, Control, 或 Computer, Information) 等因素的影响. 在一定的程度上能够反映出作战双方的效能, 甚至战争的结果. 实际中, 如果已知作战双方的各兵种的实力和相应的效能指标等, 则由该模型可对战争的发展做出评估预测. 该模型被广泛地应用于研究分析战争的重要定量工具, 也成为现代作战模拟的基本工具之一.

3.4 SARS 传播问题^①

3.4.1 问题的提出

SARS (Severe Acute Respiratory Syndrome, 严重急性呼吸道综合症, 俗称: 非典型肺炎) 是 21 世纪第一个在世界范围内传播的传染病, SARS 的爆发和蔓延给部分国家和地区的经济发展和人民生活带来了很大影响, 人们从中得到了许多重要的经验和教训, 认识到定量地研究传染病的传播规律、为预测和控制传染病蔓延创造条件的重要性. 请你对 SARS 的传播建立数学模型, 要求说明怎样才能建立一个真正能够预测以及能为预防和控制提供可靠、足够的信息的模型, 这样做的困难在哪里? 并对疫情传播所造成的影响做出估计.

^① 本案例选自 2003 年中国大学生数学建模竞赛 A 题, 根据信息工程大学信息工程学院张启慧, 芦斌, 扈红超的获奖论文整理.

3.4.2 问题的分析

实际中, SARS 的传染过程为

易感人群→病毒潜伏人群→发病人群→退出者(包括死亡者和治愈者)。

通过分析各类人群之间的转化关系, 可以建立微分方程模型来刻画 SARS 传染规律。

疫情主要受日接触率 $\lambda(t)$ 影响, 不同的时段, $\lambda(t)$ 的影响因素不同。在 SARS 传播过程中, 卫生部门的控制预防措施起着较大的作用。以采取控制措施的時刻 t_0 作为分割点, 将 SARS 传播过程分为控前和控后两个阶段。

在控前阶段, SARS 按自然传播规律传播, $\lambda(t)$ 可视为常量; 同时, 在疫情初期, 人们的防范意识比较弱, 再加上 SARS 自身的传播特点, 在个别地区出现了“超级传染事件”(SSE), 即 SARS 病毒感染者在社会上的超级传播事件。到了中后期, 随着人们防范意识的增强, SSE 发生的概率减小, 因此, SSE 在 SARS 的疫情早期对疫情的发展起到了很大的影响。SSE 其特性在于在较短的时间内, 可使传染者数目快速增加。故可将 SSE 对疫情的影响看作一个脉冲的瞬时行为, 使用脉冲微分方程描述。

控后阶段, 随着人们防范措施的增强促使日传染率 $\lambda(t)$ 减小, 引起人们防范措施增强的原因主要有两方面:

- (1) 来自于应对疫情的恐慌心理, 而迫使人们加强自身防范;
- (2) 来自于预防政策、法律法规的颁布等而加强了防范措施。

以上两者又分别受疫情数据的影响, 关系如图 3-1。



图 3-1 疫情关系图

在做定量计算时, 可以先定性分析确定各因素之间的函数关系, 再在求解过程中利用参数辨识方法确定其中的参数。

3.4.3 问题的假设与符号说明

模型的假设:

(1) 由于 SARS 的传染期不是很长, 故不考虑这段时间内的人口出生率和自然死亡率;

(2) 平均潜伏期为 6 天;

(3) 处于潜伏期的 SARS 病人不具有传染性.

符号说明:

t_0 表示从最初发现 SARS 患者到卫生部门采取预防措施的时间间隔; N 表示疫区总人口数; $S(t)$ 表示 t 时刻健康人数占总人数的比例; $I(t)$ 表示 t 时刻感染人数占总人数的比例; $E(t)$ 表示 t 时刻潜伏期的人数占总人数的比例; $Q(t)$ 表示 t 时刻退出者的人数占总人数的比例; $\lambda(t)$ 表示日接触率, 即表示每个病人平均每天有效接触的人数; $f(t)$ 表示疫情指标; $g(t)$ 表示预防措施的力度; $h(t)$ 表示人们的警惕性指标; $w(t)$ 表示防范意识; $b(t)$ 表示 t 时刻实际的新增确诊人数; $b'(t)$ 表示模型计算得到的 t 时刻新增确诊人数.

3.4.4 模型的建立

1. 各类人群的转化过程

由问题的分析, 将人群分为易感人群 S , 病毒潜伏人群 E , 发病人群 I , 退出者 Q 四类.

(1) 易感人群 S 与病毒潜伏人群 E 间的转化:

易感者和发病者有效接触后成为病毒潜伏者, 设每个发病者平均每天有效接触的易感者数为 $\lambda(t)S$, NI 个发病者平均每天能使 $\lambda(t)SNI$ 个易感者成为病毒潜伏者. 故

$$N \frac{dS}{dt} = -\lambda SNI, \text{ 即 } \frac{dS}{dt} = -\lambda SI.$$

(2) 病毒潜伏人群 E 与发病人群 I 间的转化:

病毒潜伏人群的变化等于易感人群转入的数量减去转为发病人群的数量, 即

$$\frac{dE}{dt} = S\lambda(t)I - \varepsilon E,$$

其中 ε 表示潜伏期日发病率, 根据有关文献资料^[7], 在这里取 $\varepsilon = \frac{1}{6}$.

(3) 发病人群 I 与退出者 Q 间的转化:

单位时间内退出者的变化等于发病人群的减少, 即

$$\frac{dQ}{dt} = \omega I,$$

其中 ω 表示日退出率, 根据有关资料取 $\omega = 0.008$ ^[7].

综上所述, 建立了整个系统中各类人群的转化过程, 下面将疫情传播过程分别按控前阶段和控后阶段建立相应的模型.

2. 控前阶段的自然传播模型

(1) 参数确定

日传染率 $\lambda(t)$ 在疫情的初期, SARS 按自然传播规律传播, $\lambda(t)$ 保持不变, 即为待定常数 λ_0 , 具体取值在模型求解中通过参数辨识确定.

(2) 超级传染事件(SSE)的处理

定义脉冲函数: $\delta_\varepsilon(x - x_0) = \begin{cases} \frac{1}{2\varepsilon}, & x_0 - \varepsilon < x < x_0 + \varepsilon, \\ 0, & \text{其他.} \end{cases}$

δ 函数: $\delta(x - x_0) = \lim_{\varepsilon \rightarrow 0} \delta_\varepsilon(x - x_0)$.

由问题的分析, 将 SSE 对疫情的影响看作一个瞬时的脉冲行为, 则

$$\begin{aligned} \frac{dS}{dt} &= -\lambda(t)IS - N \sum_{i=1}^m \alpha_i \delta(t - t_i), \\ \frac{dE}{dt} &= S\lambda(t)I - \varepsilon E + N \sum_{i=1}^m \alpha_i \delta(t - t_i), \end{aligned}$$

其中 m 为所加 δ 函数的个数, 在实际表现为 SSE 的个数; α_i 为第 i 个 δ 函数的强度, 根据有关统计资料, 每例 SSE 事件的平均感染人数为 20 人.

(3) 控前阶段的传播模型

综合上述的讨论, 可以得到控前阶段的自然传播模型:

$$\begin{cases} \frac{dS}{dt} = -\lambda(t)IS - N \sum_{i=1}^m \alpha_i \delta(t - t_i), \\ \frac{dE}{dt} = S\lambda(t)I - \varepsilon E + N \sum_{i=1}^m \alpha_i \delta(t - t_i), \\ \frac{dI}{dt} = \varepsilon E - \omega I, \\ \frac{dQ}{dt} = \omega I, \\ S + E + I + Q = 1, \\ S(0) = s_0, E(0) = E_0, I(0) = I_0, Q(0) = Q_0, \end{cases} \quad (3.10)$$

其中 S_0, E_0, I_0, Q_0 为系统中各类的初始值.

3. 控后阶段的传播模型

(1) 疫情指标 $f(t)$ 的确定

影响疫情指标因素主要是每日新增死亡人数 $d(t)$ 、新增确诊人数 $b(t)$ 、新增疑似病例人数 $v(t)$. 对于这三个因素归一后求加权和得到

$$f(t) = q_1 \frac{d(t)}{\max(d(t))} + q_2 \frac{b(t)}{\max(b(t))} + q_3 \frac{v(t)}{\max(v(t))},$$

其中 q_1, q_2, q_3 依次为 $d(t), b(t), v(t)$ 对疫情指标的相对影响权重, 考虑到人们对三类新增人数的敏感程度, 不妨取 $q_1 = 0.4, q_2 = 0.4, q_3 = 0.2$. 由实际统计数据知, $f(t)$ 的取值是离散的, 为此, 采用最小二乘拟合方法, 可以得到 $f(t)$ 的近似

表达式(如图 3-2). 另一方面,从离散的数据点看出,其规律大致呈韦伯分布,故可取韦伯分布密度函数

$$f(t) = \frac{m}{x_0} (t - v)^{m-1} e^{-\frac{(t-v)^m}{x_0}}.$$

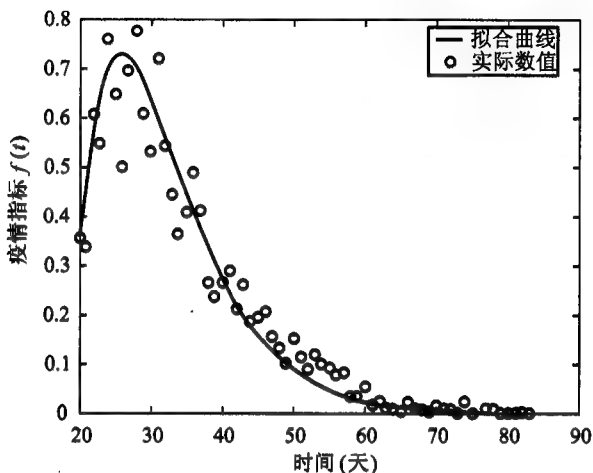


图 3-2 疫情指标的拟合曲线与实际数据比较

由参数估计可得

$$m = 2.3449, v = -1.1578, x_0 = 14.3530.$$

(2) 预防措施力度 $g(t)$ 的确定

在控后阶段,卫生部门的预防措施力度 $g(t)$ 在控制疫情的过程中起到了重要的作用,与下列因素有关:

1) 卫生部门关注的疫情来自于最近几天的疫情,不妨取近三天的疫情 $f(t)$;

2) 当 $t = t_0$ 时, $g(t)$ 有一个初始值,即为潜在的预防措施力度 k_0 ($0 < k_0 < 1$);

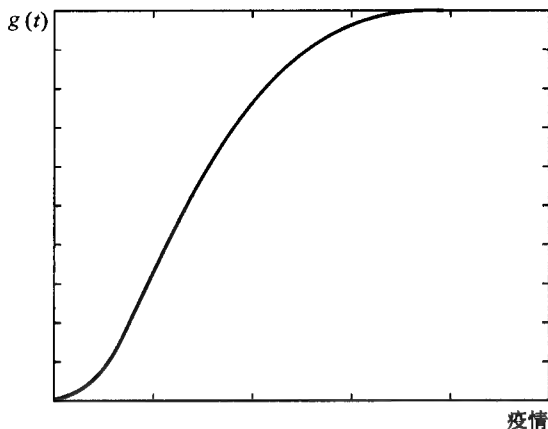
3) $g(t)$ 随疫情的增强而增加. 前期增加较为缓慢,但疫情发展到一定程度后,社会对疫情的蔓延变得敏感起来,后期预防力度加大,随之疫情指标的增长速度变慢;

4) 当疫情最严重时, $g(t)$ 最大趋向于 1.

综上所述,可以给出 $g(t)$ 随疫情变化的曲线,形态如图 3-3 所示(横坐标为疫情,纵坐标为 $g(t)$),其表达式为

$$g(t) = k_0 + k_1 \cdot (1 - e^{-\frac{f(t)^2}{\sigma_1}}),$$

其中 $k_0 + k_1 = 1$. 根据有关数据,令 $k_0 = 0.2, k_1 = 0.8$, 当 $f(t_0) = 0.58$ 时,取 $g(t_0) = 0.7$,得参数估计 $\sigma_1 = 0.1803$.

图 3-3 $g(t)$ 随 $f(t)$ 的变化

(3) 人们的警惕性指标 $h(t)$ 的确定

人们对 SARS 的警惕性程度也随疫情的变化而变化. 在公布疫情初期, 疫情的变化引起人们很大的关注, 警惕性程度随疫情的微小变化波动很大; 到中后期, 波动逐渐变缓, 直至平稳. 可用 $h(t) = k_2 - k_3 e^{-f(t)}$ 来定量刻画 $h(t)$ 与 $f(t)$ 的关系.

当 $f(t) = 0$ 时, $h(t) = 0.2$ (即为人们固有的警惕性指标); 当 $f(t) \rightarrow +\infty$ 时, $h(t) \rightarrow 1$, 参数估计得 $k_2 = 1, k_3 = 0.8$.

(4) 防范措施 $w(t)$ 的确定

由问题分析, 人们的防范措施 $w(t)$ 受预防措施力度 $g(t)$ 和警惕性指标 $h(t)$ 的影响, $g(t), h(t)$ 对 $w(t)$ 的影响作用大致相当, 可取 $w(t) = 0.5g(t) + 0.5h(t)$.

(5) 防范措施 $w(t)$ 与日传染率 $\lambda(t)$ 的关系

$\lambda(t)$ 表示发病者平均每天有效接触的人数, 由问题分析知, $\lambda(t)$ 是防范措施 $w(t)$ 的函数, 且应满足

- 1) 当防范措施 $w(t)$ 为零时, 则 $\lambda(t)$ 取最大值——控前阶段的日接触率;
- 2) 随 $w(t)$ 的增大, $\lambda(t)$ 会减小, 当 $w(t)$ 不强时, 对 $\lambda(t)$ 的变化所起的作用较小; 当 $w(t)$ 超过一定的数值时, 则对 $\lambda(t)$ 的影响效果较明显;
- 3) 当 $w(t)$ 趋近于 1 (不可能为 1) 时, 则 $\lambda(t)$ 趋近于 0.

由以上三点可以确定 $\lambda(t)$ 随 $w(t)$ 变化的曲线形态, 采用函数

$$\lambda(t) = k_4 \left(1 - e^{-\frac{(1-w(t))^2}{\sigma_2}} \right)$$

刻画此形态. 其中 σ_2 为待定常数.

(6) 控后阶段的模型

综上所述,控后阶段的 SARS 疫情的传播模型为

$$\begin{cases} \frac{dS}{dt} = -\lambda(t)IS, \\ \frac{dE}{dt} = S\lambda(t)I - \varepsilon E, \\ \frac{dI}{dt} = \varepsilon E - \omega I, \\ \frac{dQ}{dt} = \omega I, \\ \lambda(t) = k_4(1 - e^{-\frac{(1-w(t))^2}{\sigma_2}}), \\ S + E + I + Q = 1, \\ S(0) = S_0, E(0) = E_0, \\ I(0) = I_0, Q(0) = Q_0, \end{cases} \quad (3.11)$$

3.4.5 模型的求解

由于模型(3.10),(3.11)较为复杂,要求解析解是困难的,故将微分方程模型转化为差分方程求解.

以 12 例 SARS 患者作为疫情初始值,即 $I(0) = 12/N = 0.12 \times 10^{-5}$, $E(0) = 0$, $Q(0) = 0$. 求解可得实际数据 $b(t)$ 与计算结果 $b'(t)$ 的比较,如图 3-4 所示.

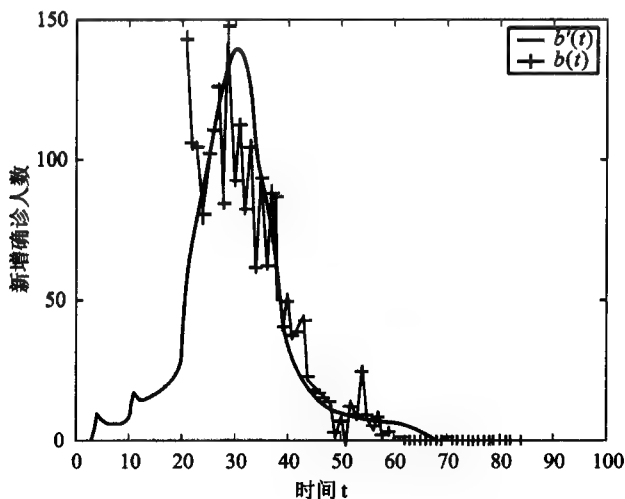


图 3-4 新增确诊人数的实际值与计算值的比较

由图可以看出 $b(t)$ 与 $b'(t)$ 的走势大致相同,且值相差不大,其中开始的小高峰是 SSE 事件造成的. 由参数辨识可以得到模型中未确定的两个待定参数 λ_0 .

$$=0.374, \sigma_2 = 28.0891.$$

3.4.6 模型的结果分析

1. 采取严格隔离措施早晚的影响

根据有关数据,对于提前5天或延后5天采取严格的预防措施的情况比较如图3-5.

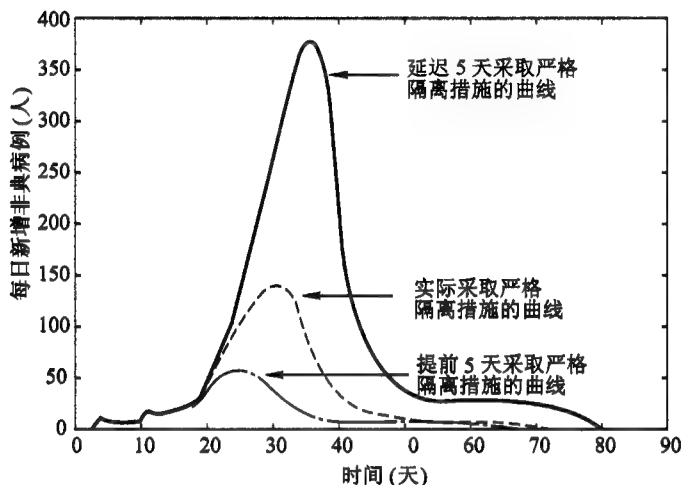


图3-5 隔离措施早晚对疫情的影响

如果卫生部门延后5天采取严格预防措施,则日新增病例峰值为376例,如果提前5天,则日新增病例峰值为55例.由此可见日新增病例的峰值对采取严格预防措施的早晚十分敏感,采取的措施越晚,疫情峰值越高,疫情周期越长.这对于指导SARS的防治工作具有重要意义,卫生部门应该在实际工作中“早发现早隔离”,采取有效的隔离预防措施.

2. 采取措施的力度对疫情的影响

预防措施力度 $g(t)$ 反映了卫生部门针对疫情所采取干预的力度,在这里我们分别取 $g(t_0) = 0.9, 0.7, 0.5$, 代入模型中,计算结果如图3-6.

从图中可看出,预防措施力度越弱,曲线的拖尾越长,甚至会再次出现疫情小高峰的现象.当 $g(t_0) = 0.5$ 时,曲线出现了第二次峰值,这表示如果在疫情刚有所下降时,就放松预防力度,疫情将会出现反弹,引起第二次疫情峰值.因此预防措施力度一定要持续,不能看到疫情有所缓和就放松警惕.

3. 人们警惕性程度对疫情的影响

对于突发性事件人们有个固有的警惕性程度,对该固有警惕性程度取 $h(t) = 0.1, 0.2, 0.3$ 代入模型求解计算得结果如图3-7.从图中可看出,固有警惕性

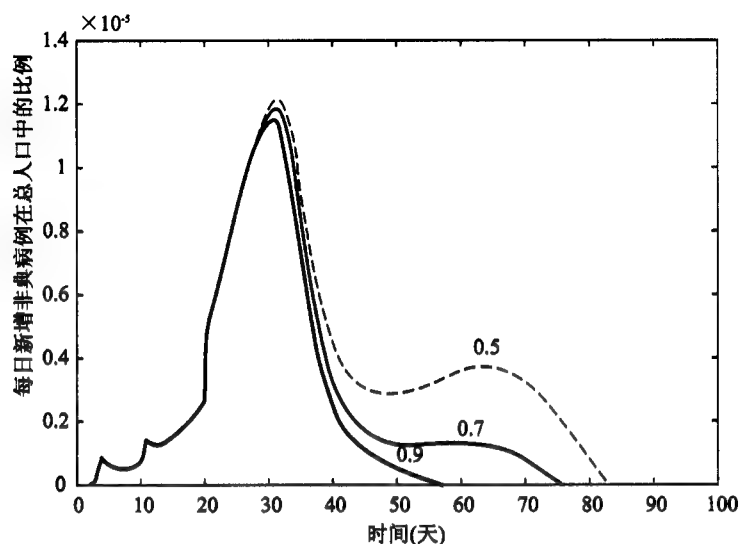


图 3-6 预防措施力度对疫情的影响

程度越小,疫情曲线拖尾越长,甚至会发生二次高峰现象,图中给出了当警惕程度为 0.1 时,就出现了二次疫情高峰。因此,卫生部门应号召群众要戒除陋习,改变生活习惯,提高防范意识,就是为了使固有警惕性程度增强,这样不仅可以使疫情不出现二次峰值,而且可以使疫情周期缩短,这也说明卫生部门加强该项措施对控制疫情是非常有效的。

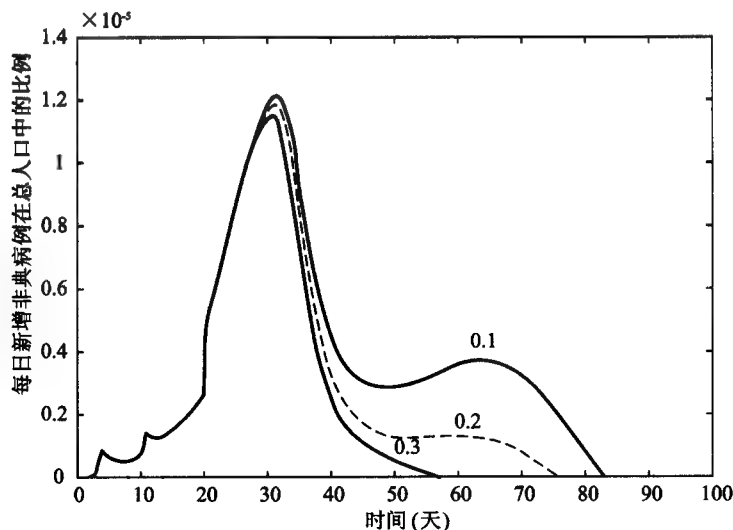


图 3-7 警惕性程度对疫情的影响

3.5 参考案例与参考文献

1. 参考案例

- (1) 药物在体内的分布与排除问题——文献[1]:132—137
- (2) 传染病问题——文献[2]:120—147
- (3) 地中海鲨鱼问题——文献[4]:92—97
- (4) 作战问题——文献[5]:122—144
- (5) 人口预测与控制问题——文献[6]:27—301

2. 参考文献

- [1] 姜启源. 数学模型. 第二版. 北京:高等教育出版社,1993
- [2] 寿纪麟. 数学建模——方法与范例. 西安:西安交通大学出版社,1993
- [3] 王柔怀等. 常微分方程讲义. 北京:人民教育出版社,1978
- [4] 赵静,但琦等. 数学建模与数学实验. 北京:高等教育出版社,2002
- [5] WILLIAM F. LUCAS. 微分方程模型. 朱煜民等译. 长沙:国防科技大学出版社,1988
- [6] 谭永基等. 数学模型. 上海:复旦大学出版社,1997
- [7] 杨方廷等. 北京 SARS 疫情过程的仿真分析. 系统仿真学报,2003,15(7):991—998

第四章 差分方程方法

在实际中,许多问题所研究的变量都是离散的形式,所建立的数学模型也是离散的,譬如,像政治、经济和社会等领域中的实际问题.有些时候,即使所建立的数学模型是连续形式,例如像常见的微分方程模型、积分方程模型等等,但是,往往都需要用计算机求数值解.这就需要将连续变量在一定的条件下进行离散化,从而将连续型模型转化为离散型模型,因此,最后都归结为求解离散形式的差分方程解的问题.关于差分方程理论和求解方法在数学建模和解决实际问题的过程中起着重要作用.

下面就不同类型的差分方程进行讨论.所谓的差分方程是指:对于一个数列 $\{x_n\}$,把数列中的前 $n+1$ 项 x_i ($i=0,1,2,\cdots,n$) 关联起来所得到的方程.

4.1 常系数线性差分方程

4.1.1 常系数线性齐次差分方程

常系数线性齐次差分方程的一般形式为

$$x_n + a_1 x_{n-1} + a_2 x_{n-2} + \cdots + a_k x_{n-k} = 0, \quad (4.1)$$

其中 k 为差分方程的阶数, a_i ($i=1,2,\cdots,k$) 为差分方程的系数,且 $a_k \neq 0$ ($k \leq n$). 对应的代数方程

$$\lambda^k + a_1 \lambda^{k-1} + a_2 \lambda^{k-2} + \cdots + a_k = 0 \quad (4.2)$$

称为差分方程(4.1)的特征方程,其特征方程的根称为特征根.

常系数线性齐次差分方程的解主要是由相应的特征根的不同情况有不同的形式.下面分别就特征根为单根、重根和复根的情况给出差分方程解的形式.

1. 特征根为单根

设差分方程(4.1)有 k 个单特征根 $\lambda_1, \lambda_2, \cdots, \lambda_k$, 则差分方程(4.1)的通解为

$$x_n = c_1 \lambda_1^n + c_2 \lambda_2^n + \cdots + c_k \lambda_k^n,$$

其中 c_1, c_2, \cdots, c_k 为任意常数,且当给定初始条件

$$x_i = x_i^{(0)} \quad (i=1,2,\cdots,k) \quad (4.3)$$

时,可以惟一确定一个特解.

2. 特征根为重根

设差分方程(4.1)有 l 个相异的特征根 $\lambda_1, \lambda_2, \dots, \lambda_l (1 \leq l < k)$, 重数分别为

m_1, m_2, \dots, m_l , 且 $\sum_{i=1}^l m_i = k$, 则差分方程(4.1)的通解为

$$x_n = \sum_{i=1}^{m_1} c_{1i} n^{i-1} \lambda_1^n + \sum_{i=1}^{m_2} c_{2i} n^{i-1} \lambda_2^n + \dots + \sum_{i=1}^{m_l} c_{li} n^{i-1} \lambda_l^n.$$

同样的,由给定的初始条件(4.3)可以惟一确定一个特解.

3. 特征根为复根

设差分方程(4.1)有一对共轭复根 $\lambda_1, \lambda_2 = \alpha \pm i\beta$ 和相异的 $k-2$ 个单根 $\lambda_3, \lambda_4, \dots, \lambda_k$, 则差分方程的通解为

$$x_n = c_1 \rho^n \cos n\theta + c_2 \rho^n \sin n\theta + c_3 \lambda_3^n + c_4 \lambda_4^n + \dots + c_k \lambda_k^n,$$

其中 $\rho = \sqrt{\alpha^2 + \beta^2}$, $\theta = \arctan \frac{\beta}{\alpha}$.

同样由给定的初始条件(4.3)可以惟一确定一个特解.

另外,对于有多个共轭复根和相异实根,或共轭复根和重根的情况,都可以类似地给出差分方程解的形式.

4.1.2 常系数线性非齐次差分方程

常系数线性非齐次差分方程的一般形式为

$$x_n + a_1 x_{n-1} + a_2 x_{n-2} + \dots + a_k x_{n-k} = f(n), \quad (4.4)$$

其中 k 为差分方程的阶数, $a_i (i=1, 2, \dots, k)$ 为差分方程的系数, $a_k \neq 0 (k \leq n)$, $f(n)$ 为已知函数.

在差分方程(4.4)中,令 $f(n) = 0$, 所得方程

$$x_n + a_1 x_{n-1} + a_2 x_{n-2} + \dots + a_k x_{n-k} = 0 \quad (4.5)$$

称为非齐次差分方程(4.4)对应的齐次差分方程,即与差分方程(4.1)的形式相同.

求解非齐次差分方程通解的一般方法为

首先求对应的齐次差分方程(4.5)的通解 x_n^* , 然后求非齐次差分方程(4.4)的一个特解 $x_n^{(0)}$, 则

$$x_n = x_n^* + x_n^{(0)}$$

为非齐次差分方程(4.4)的通解.

关于求 x_n^* 的方法同求差分方程(4.1)的方法相同. 对于求非齐次方程(4.4)的特解 $x_n^{(0)}$ 的方法,可以用观察法确定,也可以根据 $f(n)$ 的特性用待定系数法确定,具体方法可参照常系数线性非齐次微分方程求特解的方法.

4.2 差分方程的平衡点及其稳定性

一般说来,差分方程的求解是困难的,实际中往往不要求出差分方程的一般解,而只需要研究它的平衡点及其稳定性即可.

4.2.1 一阶线性常系数差分方程

一阶线性常系数差分方程的一般形式为

$$x_{k+1} + ax_k = b, k = 0, 1, 2, \dots,$$

其中 a, b 为常数, 它的平衡点由代数方程 $x + ax = b$ 求解得到, 不妨记为 x^* .

如果 $\lim_{k \rightarrow \infty} x_k = x^*$, 则称平衡点 x^* 是稳定的, 否则是不稳定的.

为了便于研究平衡点 x^* 的稳定性问题, 一般将其转化为求方程 $x_{k+1} + ax_k = 0$ 的平衡点 $x^* = 0$ 的稳定性问题. 事实上, 由

$$x_{k+1} + ax_k = 0$$

可以解得

$$x_k = (-a)^k x_0,$$

于是 $x^* = 0$ 是稳定的平衡点的充要条件是: $|a| < 1$.

4.2.2 一阶线性常系数差分方程组

一阶线性常系数齐次差分方程组的一般形式为

$$x(k+1) + Ax(k) = 0, k = 0, 1, 2, \dots,$$

其中 $x(k)$ 为 n 维向量, A 为 $n \times n$ 阶常数矩阵.

它的平衡点 $x^* = 0$ 是稳定的充要条件是 A 的所有特征根都有 $|\lambda_i| < 1$ ($i = 1, 2, \dots, n$).

对于一阶线性常系数非齐次差分方程组

$$x(k+1) + Ax(k) = B, k = 0, 1, 2, \dots$$

的情况同样给出.

4.2.3 二阶线性常系数差分方程

二阶线性常系数齐次差分方程的一般形式为

$$x_{k+2} + a_1 x_{k+1} + a_2 x_k = 0, k = 0, 1, 2, \dots,$$

其中 a_1, a_2 为常数, 其平衡点 $x^* = 0$ 是稳定的充要条件是特征方程 $\lambda^2 + a_1 \lambda + a_2 = 0$ 的根 λ_1, λ_2 满足 $|\lambda_1| < 1, |\lambda_2| < 1$.

对于一般的 $x_{k+2} + a_1 x_{k+1} + a_2 x_k = b$ 的平衡点的稳定性问题同样给出. 类似

地,也可直接推广到 n 阶线性差分方程的情况.

4.2.4 一阶非线性差分方程

一阶非线性差分方程的一般形式为

$$x_{k+1} = f(x_k), k = 0, 1, 2, \dots,$$

其中 f 为已知函数,其平衡点定义为方程 $x = f(x)$ 的解 x^* .

事实上,将 $f(x_k)$ 在 x^* 处作一阶的泰勒展开有

$$x_{k+1} \approx f'(x^*)(x_k - x^*) + f(x^*),$$

则 x^* 也是一阶线性差分方程 $x_{k+1} = f'(x^*)(x_k - x^*) + f(x^*)$ 的平衡点,故此,平衡点 x^* 稳定的充要条件是 $|f'(x^*)| < 1$.

4.3 连续模型的差分方法

4.3.1 微分的差分方法

已知 $f(x)$ 在点 x_k 处的函数值 $f(x_k)$ ($k = 0, 1, \dots, n+1$),且 $a = x_0 < x_1 < \dots < x_{n+1} = b$,试求函数的导数值 $f'(x_k)$ ($k = 1, 2, \dots, n$).

根据导数的定义,用差商代替微商,则有下列的差分公式.

向前差:

$$f'(x_k) \approx \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k} (k = 1, 2, \dots, n),$$

向后差:

$$f'(x_k) \approx \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} (k = 1, 2, \dots, n);$$

中心差:

$$f'(x_k) \approx \frac{f(x_{k+1}) - f(x_{k-1}))}{x_{k+1} - x_{k-1}} (k = 1, 2, \dots, n).$$

4.3.2 定积分的差分方法

已知函数 $f(x)$ 在点 x_k 处的函数值 $f(x_k)$ ($k = 0, 1, \dots, n$),且在 $[a, b]$ 上可积,试求函数在 $[a, b]$ 上的积分值 $\int_a^b f(x) dx$.

根据定积分的定义,则有一般的求积公式

$$\int_a^b f(x) dx \approx \sum_{k=0}^n A_k f(x_k),$$

其中 A_k 为求积系数, 它与 x_k 的选取方法有关. 取不同的求积系数, 可以得不同的求积公式.

对于等距节点 $x_k = a + kh$ ($k = 0, 1, \dots, n$), 其中步长 $h = \frac{b-a}{n}$ 为很小的数,

则有如下的求积公式.

(1) 复化矩形公式:

$$\int_a^b f(x) dx \approx h \sum_{k=0}^{n-1} f\left[a + \left(k + \frac{1}{2}\right)h\right].$$

(2) 复化梯形公式:

$$\begin{aligned} \int_a^b f(x) dx &\approx \frac{h}{2} \sum_{k=0}^{n-1} [f(x_k) + f(x_{k+1})] \\ &= \frac{h}{2} \left[f(a) + 2 \sum_{k=1}^{n-1} f(x_k) + f(b) \right]. \end{aligned}$$

(3) 复化辛普森 (Simpson) 公式:

$$\begin{aligned} \int_a^b f(x) dx &\approx \frac{h}{6} \sum_{k=0}^{n-1} \left[f(x_k) + 4f\left(x_{k+\frac{1}{2}}\right) + f(x_{k+1}) \right] \\ &= \frac{h}{6} \left[f(a) + 4 \sum_{k=0}^{n-1} f\left(x_{k+\frac{1}{2}}\right) + 2 \sum_{k=1}^{n-1} f(x_k) + f(b) \right], \end{aligned}$$

其中 $x_{k+\frac{1}{2}} = \frac{1}{2}(x_k + x_{k+1})$ 为子区间 $[x_k, x_{k+1}]$ ($k = 0, 1, \dots, n-1$) 的中点.

(4) 复化柯特斯 (Cotes) 公式:

$$\begin{aligned} \int_a^b f(x) dx &\approx \frac{h}{90} \left[7f(a) + 32 \sum_{k=0}^{n-1} f\left(x_{k+\frac{1}{4}}\right) + 12 \sum_{k=0}^{n-1} f\left(x_{k+\frac{1}{2}}\right) \right. \\ &\quad \left. + 32 \sum_{k=0}^{n-1} f\left(x_{k+\frac{3}{4}}\right) + 12 \sum_{k=1}^{n-1} f(x_k) + 7f(b) \right], \end{aligned}$$

其中 $x_{k+\frac{1}{4}}, x_{k+\frac{1}{2}}, x_{k+\frac{3}{4}}$ 为子区间 $[x_k, x_{k+1}]$ ($k = 0, 1, \dots, n-1$) 中的四等分点.

4.3.3 常微分方程的差分方法

1. 一阶常微分方程的差分方法

设一阶常微分方程的定解问题为

$$\begin{cases} y' = f(x, y), \\ y(x_0) = y_0, \end{cases} \quad (4.6)$$

其中函数 $f(x, y)$ 关于 y 满足李普希兹条件, 即保证问题 (4.6) 解的存在惟一性.

现在的问题是求方程 (4.6) 在一系列节点 $x_1 < x_2 < \dots < x_n < \dots$ 处的近似数值解 $y_1, y_2, \dots, y_n, \dots$. 不妨假设步长为 $h = x_{n+1} - x_n$ 为常数. 在此, 我们根据微分的差分方法, 即用差商来近似代替微商, 再利用“步进式”方法, 可以给出求解问

题(4.6)的差分方法.

(1) 单步欧拉(Euler)公式

用差商 $\frac{y(x_{n+1}) - y(x_n)}{h}$ 近似代替 $y'(x_n) = f(x_n, y(x_n))$ 中的导数, 则可得差分公式

$$y_{n+1} = y_n + hf(x_n, y_n), \quad n = 0, 1, 2, \dots,$$

其精度为 $O(h^2)$ 阶的.

(2) 两步欧拉公式

用差商 $\frac{y(x_{n+1}) - y(x_{n-1}))}{2h}$ 近似代替 $y'(x_n) = f(x_n, y(x_n))$ 中的导数, 则可得差分公式

$$y_{n+1} = y_{n-1} + 2hf(x_n, y_n), \quad n = 1, 2, \dots.$$

两步法需要用到前两步的信息, 一般不能自行起步, 需先用单步法求出 y_1 . 其精度是 $O(h^3)$ 阶的.

(3) 梯形公式

对于方程 $y' = f(x, y)$ 的两边在 $[x_n, x_{n+1}]$ 上求积分得

$$y(x_{n+1}) = y(x_n) + \int_{x_n}^{x_{n+1}} f(x, y(x)) dx.$$

利用积分的差分方法中梯形公式求解积分

$$\int_{x_n}^{x_{n+1}} f(x, y(x)) dx \approx \frac{h}{2} [f(x_n, y(x_n)) + f(x_{n+1}, y(x_{n+1}))],$$

则

$$y(x_{n+1}) \approx y(x_n) + \frac{h}{2} [f(x_n, y(x_n)) + f(x_{n+1}, y(x_{n+1}))].$$

离散化即可得到微分方程的梯形差分公式

$$y_{n+1} = y_n + \frac{h}{2} [f(x_n, y_n) + f(x_{n+1}, y_{n+1})], \quad n = 0, 1, 2, \dots.$$

这是一个隐式格式, 计算量较大, 一般不单独使用. 其精度也是 $O(h^3)$ 阶的.

(4) 改进的欧拉公式

由于单步欧拉公式精度低, 但计算量小; 矩形公式精度高, 但计算量大, 为此我们综合运用这两种方法就可以得到改进的欧拉公式, 其精度为 $O(h^3)$ 阶的.

预报: $\bar{y}_{n+1} = y_n + hf(x_n, y_n), \quad n = 0, 1, \dots.$

校正: $y_{n+1} = y_n + \frac{h}{2} [f(x_n, y_n) + f(x_{n+1}, \bar{y}_{n+1})], \quad n = 0, 1, \dots$

或写成平均化形式:

$$\begin{cases} y_p = y_n + hf(x_n, y_n), \\ y_c = y_n + hf(x_{n+1}, y_p), \\ y_{n+1} = \frac{1}{2}(y_p + y_c), \quad n = 0, 1, 2, \dots \end{cases}$$

(5) 龙格-库塔(Runge-Kutta)法

龙格-库塔方法的基本思想:

对于微分方程的定解问题(4.6), 考虑差商 $\frac{y(x_{n+1}) - y(x_n)}{h}$, 根据拉格朗日

(Lagrange)微分中值定理可得

$$y(x_{n+1}) - y(x_n) = hy'(\xi) = hf(\xi, y(\xi)), \quad x_n < \xi < x_{n+1},$$

记 $Y^* = f(\xi, y(\xi))$, 称为 $[x_n, x_{n+1}]$ 上的平均变化率, 则 $y(x_{n+1}) = y(x_n) + hY^*$.

现在的问题只要寻找一种计算 Y^* 的方法.

如果取 $Y^* \approx f(x_n, y_n) = Y_1$, 则就是欧拉公式.

如果取 $Y^* \approx \frac{1}{2}[f(x_n, y_n) + f(x_{n+1}, y_{n+1})] = Y_2$, 则相应的就是改进的欧拉公式.

现在, 我们取 m 个点 $(x_n + \alpha_i h, y_n + \beta_i h) \in [x_n, x_{n+1}]$ ($i = 1, 2, \dots, m$), 用 f 在这 m 个点的函数值的加权平均作为 Y^* 的近似值, 即

$$Y^* \approx \sum_{i=1}^m w_i f(x_n + \alpha_i h, y_n + \beta_i h),$$

其中 w_i 为权系数. 则有

$$y_{n+1} = y_n + h \sum_{i=1}^m w_i f(x_n + \alpha_i h, y_n + \beta_i h), \quad (4.7)$$

其中 α_i, β_i, w_i 为待定系数.

实际中, 适当选择 α_i, β_i, w_i , 使得公式有更高的精度, 这就是龙格-库塔方法的思想.

二阶龙格-库塔公式:

在 $[x_n, x_{n+1}]$ 内取中点 $x_{n+\frac{1}{2}} = x_n + \frac{1}{2}h$, 则可取 $w_1 = 0, w_2 = 1, \alpha = \beta = \frac{1}{2}$ 代入

(4.7) 式得到二阶龙格-库塔公式, 其精度为 $O(h^3)$ 阶.

$$\begin{cases} y_{n+1} = y_n + hY_2, \\ Y_1 = f(x_n, y_n), \\ Y_2 = f\left(x_n + \frac{h}{2}, y_n + \frac{h}{2}Y_1\right), \end{cases} \quad n = 0, 1, 2, \dots$$

三阶龙格-库塔公式:

在 $[x_n, x_{n+1}]$ 内任取二点 $x_{n+p} = x_n + ph, x_{n+q} = x_n + qh (0 < p < q \leq 1)$, 类似的方法可以得到三阶的龙格-库塔公式

$$\begin{cases} y_{n+1} = y_n + \frac{h}{6}(Y_1 + 4Y_2 + Y_3), \\ Y_1 = f(x_n, y_n), \\ Y_2 = f\left(x_n + \frac{h}{2}, y_n + \frac{h}{2}Y_1\right), \\ Y_3 = f\left(x_n + h, y_n + h(-Y_1 + 2Y_2)\right), \end{cases} \quad n = 0, 1, 2, \dots,$$

其精度是 $O(h^4)$ 阶的, 常用的是三阶的情况.

四阶龙格-库塔公式:

类似的方法可以得到四阶龙格-库塔公式, 其精度是 $O(h^5)$ 阶的.

$$\begin{cases} y_{n+1} = y_n + \frac{h}{6}(Y_1 + 2Y_2 + 2Y_3 + Y_4), \\ Y_1 = f(x_n, y_n), \\ Y_2 = f\left(x_n + \frac{h}{2}, y_n + \frac{h}{2}Y_1\right), \\ Y_3 = f\left(x_n + \frac{h}{2}, y_n + \frac{h}{2}Y_2\right), \\ Y_4 = f(x_{n+1}, y_n + hY_3), \end{cases} \quad n = 0, 1, 2, \dots$$

2. 一阶常微分方程组的差分方法

将前面的单个方程中的变量和函数视为向量, 相应的差分方法即可用于由多个方程组成的一阶方程组的情形.

对于二个方程的方程组

$$\begin{cases} y' = f(x, y, z), y(x_0) = y_0, \\ z' = g(x, y, z), z(x_0) = z_0, \end{cases} \quad (4.8)$$

设以 y_n, z_n 表示函数在节点 $x_n = x_0 + nh, n = 1, 2, \dots$ 上的近似解, 则有改进的欧拉公式:

$$\begin{aligned} \text{预报: } & \begin{cases} \bar{y}_{n+1} = y_n + hf(x_n, y_n, z_n), \\ \bar{z}_{n+1} = z_n + hg(x_n, y_n, z_n), \end{cases} \\ \text{校正: } & \begin{cases} y_{n+1} = y_n + \frac{h}{2}[f(x_n, y_n, z_n) + f(x_{n+1}, \bar{y}_{n+1}, \bar{z}_{n+1})], \\ z_{n+1} = z_n + \frac{h}{2}[g(x_n, y_n, z_n) + g(x_{n+1}, \bar{y}_{n+1}, \bar{z}_{n+1})], \end{cases} \end{aligned}$$

四阶龙格-库塔公式:

$$\begin{cases} y_{n+1} = y_n + \frac{h}{6}(Y_1 + 2Y_2 + 2Y_3 + Y_4), \\ z_{n+1} = z_n + \frac{h}{6}(Z_1 + 2Z_2 + 2Z_3 + Z_4), \end{cases} \quad n = 0, 1, 2, \dots,$$

其中

$$\begin{cases} Y_1 = f(x_n, y_n, z_n), \\ Z_1 = g(x_n, y_n, z_n), \\ Y_2 = f\left(x_n + \frac{h}{2}, y_n + \frac{h}{2}Y_1, z_n + \frac{h}{2}Z_1\right), \\ Z_2 = g\left(x_n + \frac{h}{2}, y_n + \frac{h}{2}Y_1, z_n + \frac{h}{2}Z_1\right), \\ Y_3 = f\left(x_n + \frac{h}{2}, y_n + \frac{h}{2}Y_2, z_n + \frac{h}{2}Z_2\right), \\ Z_3 = g\left(x_n + \frac{h}{2}, y_n + \frac{h}{2}Y_2, z_n + \frac{h}{2}Z_2\right), \\ Y_4 = f(x_{n+1}, y_n + hY_3, z_n + hZ_3), \\ Z_4 = g(x_{n+1}, y_n + hY_3, z_n + hZ_3), \end{cases} \quad n = 0, 1, 2, \dots.$$

其他的公式也都可以类似得到,即相当于同时求解多个一阶方程,从方法上没有本质的差别.

3. 高阶常微分方程的差分方法

对于某些高阶方法的定解问题,原则上可以转化为一阶方程组来求解.譬如,对于如下的二阶微分方程的定解问题

$$\begin{cases} y'' = f(x, y, y'), \\ y(x_0) = y_0, y'(x_0) = y'_0. \end{cases}$$

若令 $z = y'$, 则可化为一阶方程组的定解问题

$$\begin{cases} z' = f(x, y, z), \\ y' = z, \\ y(x_0) = y_0, z(x_0) = y'_0. \end{cases} \quad (4.9)$$

实际上, (4.9) 式可以视为 (4.8) 式的特例, 类似地可以得到相应的求解差分公式.

4.4 最优捕鱼问题^①

4.4.1 问题的提出

假设鳊鱼可分为 4 个年龄组: 称 1、2、3、4 龄鱼. 各年龄组每条鱼的平均重量分别为 5.07, 11.55, 17.86, 22.99 (g); 各年龄组鱼的自然死亡率均为 0.8 (1/年); 这种鱼为季节性集中产卵繁殖, 产卵孵化期为每年的最后 4 个月, 平均每条 4 龄鱼的产卵量为 1.109×10^5 (个), 3 龄鱼的产卵量为这个数的一半, 2 龄和 1 龄鱼不产卵. 卵孵化并成活为 1 龄鱼, 成活率 (1 龄鱼条数与产卵量 n 之比) 为 $\frac{1.22 \times 10^{11}}{1.22 \times 10^{11} + n}$.

渔业部门规定, 每年只允许在产卵孵化期前的 8 个月内进行捕捞作业. 如果每年投入的捕捞能力固定不变, 即固定努力量捕捞, 这时单位时间捕捞量将与各年龄组鱼群条数成正比, 比例系数称为捕捞强度系数. 通常使用 12 个月捕捞量

续表

| 日期 | 7.5 | | 7.6 | | 7.7 | | 7.8 | | 7.9 | | 7.10 | |
|-----------|---------|---------|---------|---------|---------|---------|--------|--------|--------|--------|-------|-------|
| 时间 t_i | 8:00 | 20:00 | 8:00 | 20:00 | 8:00 | 20:00 | 8:00 | 20:00 | 8:00 | 20:00 | 8:00 | 20:00 |
| 节点 x_i | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
| 排沙量 y_i | 306 800 | 300 000 | 271 400 | 231 000 | 160 000 | 111 000 | 91 000 | 54 000 | 45 500 | 30 000 | 8 000 | 4 500 |

函数 $y = S(x)$ 所满足的条件为

$$(1) S(x_i) = y_i, \quad i = 0, 1, 2, \dots, 23;$$

$$(2) y'_0 = S'(x_0) \approx \frac{y_2 - y_1}{x_2 - x_1} = 56\,400, \quad y'_{23} = S'(x_{23}) \approx \frac{y_{23} - y_{22}}{x_{23} - x_{22}} = -3\,500.$$

取 $S(x)$ 的三次 B 样条函数一般形式为

$$S(x) = \sum_{j=-1}^{24} c_j \Omega_3\left(\frac{x - x_0}{h} - j\right),$$

其中 c_j ($j = -1, 0, 1, \dots, 24$) 为待定常数, $h = 1$. 在这里

$$\Omega_3(x) = \begin{cases} \frac{1}{2}|x|^3 - x^2 + \frac{2}{3}, & |x| \leq 1, \\ -\frac{1}{6}|x|^3 + x^2 - 2|x| + \frac{4}{3}, & 1 < |x| < 2, \\ 0, & |x| \geq 2, \end{cases}$$

且易知

$$\Omega_3(x) = \begin{cases} \frac{2}{3}, & x = 0, \\ \frac{1}{6}, & x = \pm 1, \\ 0, & |x| \geq 2 \end{cases} \quad \text{和} \quad \Omega'_3(x) = \begin{cases} 0, & x = 0, \\ \mp \frac{1}{2}, & x = \pm 1, \\ 0, & |x| \geq 2. \end{cases}$$

根据 B 样条函数的性质, $S''(x)$ 在 $[x_0, x_{23}]$ 上连续, 则有

$$y' = S'(x) = \sum_{j=-1}^{24} c_j \Omega'_3(x - x_0 - j)$$

由插值条件 (1), (2) 可得到下列方程组

$$\begin{cases} S(x_i) = \sum_{j=-1}^{24} c_j \Omega_3(i - j) = y_i, & i = 0, 1, 2, \dots, 23, \\ S'(x_0) = \sum_{j=-1}^{24} c_j \Omega'_3(-j) = y'_0, \\ S'(x_{23}) = \sum_{j=-1}^{24} c_j \Omega'_3(23 - j) = y'_{23}, \end{cases}$$

即

$$\begin{cases} c_{i-1} + 4c_i + c_{i+1} = 6y_i, & i = 0, 1, 2, \dots, 23, \\ -c_{-1} + c_1 = 2y'_0, \\ -c_{23} + c_{24} = 2y'_{23}. \end{cases}$$

将 $c_{-1} = c_1 - 2y'_0$, $c_{24} = c_{23} + 2y'_{23}$ 代入前 24 个方程中的第一个和最后一个, 便可得到方程组 $AC = F$, 其中

$$A = \begin{bmatrix} 4 & 2 & & & & \\ & 1 & 4 & 1 & & \\ & & 1 & 4 & 1 & \\ & & & \ddots & \ddots & \ddots \\ & & & & 1 & 4 & 1 \\ & & & & & 2 & 4 \end{bmatrix}_{24 \times 24}, \quad C = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{22} \\ c_{23} \end{bmatrix},$$

$$F = \begin{bmatrix} 6y_0 + 2y'_0 \\ 6y_1 \\ 6y_2 \\ \vdots \\ 6y_{22} \\ 6y_{23} - 2y'_{23} \end{bmatrix} = \begin{bmatrix} 458 & 400 \\ 684 & 000 \\ 945 & 000 \\ \vdots & \\ 48 & 000 \\ 34 & 000 \end{bmatrix}.$$

显然 A 为满秩阵, 方程组 $AC = F$ 一定有解, 用消元法求解可得问题的解为

$$c_0 = 56\,044.398\,3, \quad c_1 = 117\,111.203\,4, \quad c_2 = 159\,510.788\,2,$$

$$c_3 = 189\,845.643\,9, \quad c_4 = 203\,106.636\,1, \quad c_5 = 239\,727.811\,8,$$

$$c_6 = 249\,182.116\,8, \quad c_7 = 263\,543.721\,1, \quad c_8 = 287\,842.998\,9,$$

$$c_9 = 302\,284.283\,2, \quad c_{10} = 317\,419.868\,4, \quad c_{11} = 304\,836.243\,1,$$

$$c_{12} = 307\,635.159\,3, \quad c_{13} = 305\,423.119\,6, \quad c_{14} = 270\,672.362\,2,$$

$$c_{15} = 240\,287.431\,4, \quad c_{16} = 154\,177.912\,0, \quad c_{17} = 103\,000.920\,4,$$

$$c_{18} = 99\,818.406\,2, \quad c_{19} = 43\,725.454\,7, \quad c_{20} = 49\,279.775\,0,$$

$$c_{21} = 32\,155.445\,2, \quad c_{22} = 2\,098.444\,2, \quad c_{23} = 7\,450.777\,9,$$

$$c_{-1} = c_1 - 2y'_0 = 4\,311.203\,4, \quad c_{24} = c_{23} + 2y'_{23} = -450.777\,9.$$

将 c_j ($j = -1, 0, 1, \dots, 24$) 代入

$$y = S(x) = \sum_{j=-1}^{24} c_j \Omega_3(x-1-j), \quad (5.20)$$

即得排沙量的变化规律. 由(5.19)式和(5.20)式可得到第*i*时间段(12小时为一段)内,任意时刻 $t \in [0, 12]$ 的排沙量. 则总的排沙量为

$$Y = \int_1^{24} S(x) dx = \int_8^{284} \sum_{j=-1}^{24} c_j \Omega_3(t-1-j) dt,$$

经计算可得 $Y = 1.844 \times 10^{11}$ t, 即从6月29日至7月10日小浪底水库排沙总量大约为1 844 亿 t, 此与媒体报道的排沙量基本相符.

对于问题(2), 研究排沙量与水量的关系, 从试验数据可以看出, 开始排沙量是随着水流量的增加而增长, 而后是随着水流量的减少而减少. 显然, 变化规律并非是线性的关系, 为此, 把问题分为两部分, 从开始水流量增加到最大值 $2\,720 \text{ m}^3/\text{s}$ (即增长的过程)为第一阶段, 从水流量的最大值到结束为第二阶段, 分别来研究水流量与排沙量的关系. 具体数据如表5-3和5-4.

表5-3 第一阶段试验观测数据

单位: 水流为 m^3/s , 含沙量为 kg/m^3

| 序号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 水流量 x_i | 1 800 | 1 900 | 2 100 | 2 200 | 2 300 | 2 400 | 2 500 | 2 600 | 2 650 | 2 700 | 2 720 |
| 含沙量 y_i | 32 | 60 | 75 | 85 | 90 | 98 | 100 | 102 | 108 | 112 | 115 |

表5-4 第二阶段试验观测数据

单位: 水流为 m^3/s , 含沙量为 kg/m^3

| 序号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|-----------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|
| 水流量 x_i | 2 650 | 2 600 | 2 500 | 2 300 | 2 200 | 2 000 | 1 850 | 1 820 | 1 800 | 1 750 | 1 500 | 1 000 | 900 |
| 含沙量 y_i | 116 | 118 | 120 | 118 | 105 | 80 | 60 | 50 | 40 | 32 | 20 | 8 | 5 |

对于第一阶段, 由表5-3用MATLAB作图(如图5-3)可以看出其变化趋势, 我们用多项式作最小二乘拟合.

设拟合函数为

$$\phi(x) = \sum_{k=1}^m a_k x^k,$$

确定待定常数 a_k ($k=0, 1, 2, \dots, m$)使得

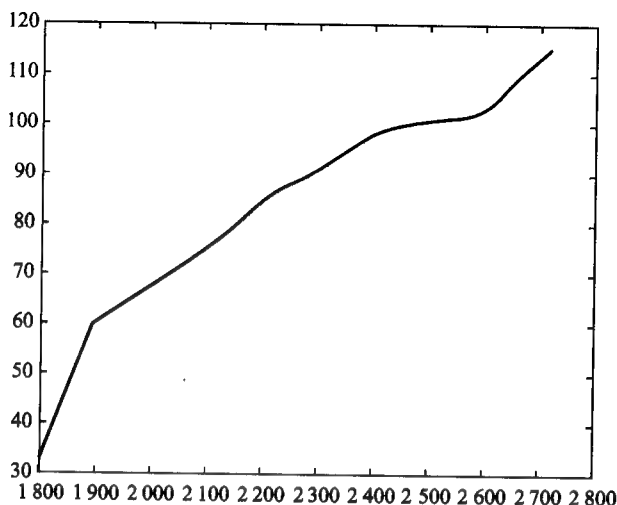


图 5-3 第一阶水流量与排沙量的关系

$$S = \sum_{i=1}^{11} [\phi(x_i) - y_i]^2 = \sum_{i=1}^{11} \left[\sum_{k=0}^m a_k x_i^k - y_i \right]^2$$

有最小值. 于是可以得到正规方程组为

$$\sum_{j=0}^m \left(\sum_{i=1}^{11} x_i^{k+j} \right) a_j = \sum_{i=1}^{11} y_i x_i^k, \quad k = 0, 1, 2, \dots, m$$

当 $m=3$ 时, 即取三次多项式拟合, 则

$$\left(\sum_{i=1}^{11} x_i^k \right) a_0 + \left(\sum_{i=1}^{11} x_i^{k+1} \right) a_1 + \left(\sum_{i=1}^{11} x_i^{k+2} \right) a_2 + \left(\sum_{i=1}^{11} x_i^{k+3} \right) a_3 = \sum_{i=1}^{11} y_i x_i^k,$$

$$k = 0, 1, 2, 3.$$

求解可得 $a_0 = -2\,492.931\,8$, $a_1 = 3.178\,4$, $a_2 = -1.317\,2 \times 10^{-3}$, $a_3 = 1.842\,3 \times 10^{-7}$.

于是可得拟合多项式为 $\phi_3(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$, 最小误差为 $S = 72.847$, 拟合效果如图 5-4.

类似地, 当 $m=4$ 时, 即取四次多项式拟合, 则正规方程组为

$$\left(\sum_{i=1}^{11} x_i^k \right) a_0 + \left(\sum_{i=1}^{11} x_i^{k+1} \right) a_1 + \left(\sum_{i=1}^{11} x_i^{k+2} \right) a_2 + \left(\sum_{i=1}^{11} x_i^{k+3} \right) a_3$$

$$+ \left(\sum_{i=1}^{11} x_i^{k+4} \right) a_4 = \sum_{i=1}^{11} y_i x_i^k, \quad k = 0, 1, 2, 3, 4.$$

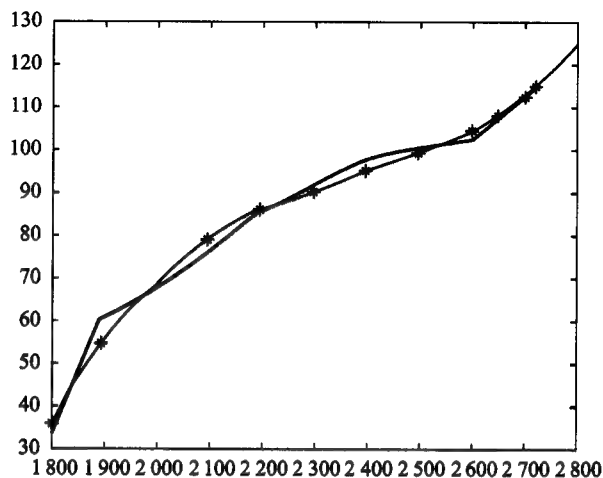


图 5-4 三次拟合效果,带 * 号的为拟合曲线

求解可得

$$a_0 = -7\,434.655\,7, a_1 = 12.062\,4, a_2 = -7.262\,6 \times 10^{-3},$$

$$a_3 = 1.94 \times 10^{-6}, a_4 = -1.931\,2 \times 10^{-10}.$$

于是可得拟合多项式为 $\phi_4(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$, 最小误差为 $S = 66.102$, 拟合效果如图 5-5.

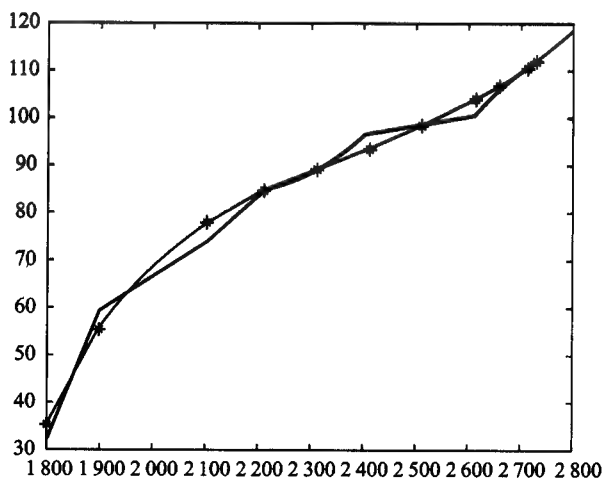


图 5-5 四次拟合效果,带 * 号的为拟合曲线

从上面的三次多项式拟合和四次多项式拟合效果来看,差别不大.基本可以看出排沙量与水流量的关系.

对于第二阶段,由表 5-4 可以类似地处理.我们用线性最小二乘法作三次和四次多项式拟合.拟合效果如图 5-6 和图 5-7,最小误差分别为 $S = 459.5$ 和 $S = 236.1$.

从拟合效果来看,显然四次多项式拟合要比三次多项式拟合好得多.

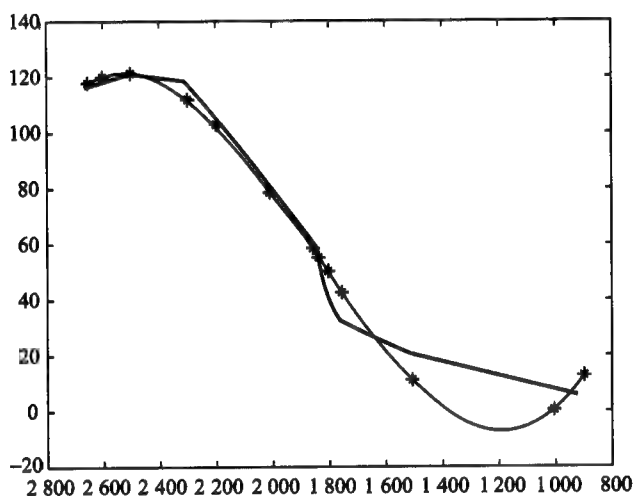


图 5-6 第二段三次多项式拟合效果

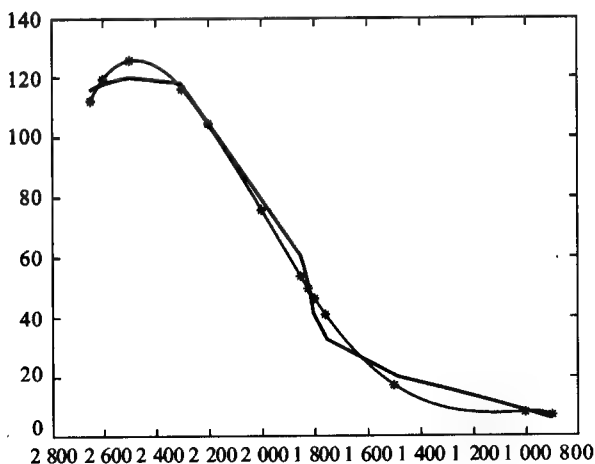


图 5-7 第二段四次多项式拟合效果

5.6 参考案例与参考文献

1. 参考案例

- (1) 农作物施肥效果问题——文献[1]:108—114
- (2) 非线性交调频率的设计问题——文献[1]:115—118
- (3) 水塔的水流量问题——文献[2]:65—70
- (4) 海底地形图测量问题——文献[2]:151—156
- (5) 万能拉拨机的凸轮设计问题——文献[3]:19—23
- (6) 血管的三维重建问题——文献[4]:40—47

2. 参考文献

- [1] 《数学模型》编写组. 数学模型. 广州:华南理工大学出版社,2001
- [2] 叶其孝. 数学建模教育与国际数学建模竞赛. 合肥:《工科数学》杂志社,1994
- [3] 樊正复. 万能拉拨机的凸轮设计. 数学的实践与认识,1989(4):19—23
- [4] 刘承平. 数学建模立法. 北京:高等教育出版社,2002
- [5] 寿纪麟. 数学建模——方法与范例. 西安:西安交通大学出版社,1993
- [6] 吴翊等. 数学建模的理论与实践. 长沙:国防科技大学出版社,2001
- [7] 王能超. 数值分析简明教程. 北京:高等教育出版社,2000
- [8] 李岳生. 样条与插值. 上海:上海科学技术出版社,1983

第六章 层次分析方法

层次分析 (Analytic Hierarchy Process, 简记 AHP) 是一种定性和定量相结合的、系统化的、层次化的分析方法. 它是将半定性、半定量问题转化为定量问题的行之有效的一种方法, 使人们的思维过程层次化. 通过逐层比较多种关联因素来为分析、决策、预测或控制事物的发展提供定量依据, 它特别适用于那些难于完全用定量进行分析的复杂问题, 为解决这类问题提供一种简便实用的方法. 因此, 它在计算、制定计划、资源分配、排序、政策分析、军事管理、冲突求解及决策预报等领域都有广泛的应用.

6.1 层次分析的一般方法

层次分析法解决问题的基本思想与人们对一个多层次、多因素、复杂的决策问题的思维过程基本一致, 最突出的特点是分层比较, 综合优化. 其解决问题的基本步骤如下:

- (1) 分析系统中各因素之间的关系, 建立系统的递阶层次结构, 一般层次结构分为三层, 第一层为目标层, 第二层为准则层, 第三层为方案层;
- (2) 构造两两比较矩阵 (判断矩阵), 对于同一层次的各因素关于上一层中某一准则 (目标) 的重要性进行两两比较, 构造出两两比较的判断矩阵;
- (3) 由比较矩阵计算被比较因素对每一准则的相对权重, 并进行判断矩阵的一致性检验;
- (4) 计算方案层对目标层的组合权重和组合一致性检验, 并进行排序.

6.1.1 层次结构图

利用层次分析法研究问题时, 首先要把与问题有关的各种因素层次化, 然后构造出一个树状结构的层次结构模型, 称为**层次结构图**. 一般问题的层次结构图分为三层, 如图 6-1.

最高层为目标层 (O): 问题决策的目标或理想结果, 只有一个元素.

中间层为准则层 (C): 包括为实现目标所涉及的中间环节各因素, 每一因素为一准则, 当准则多于 9 个时可分为若干个子层.

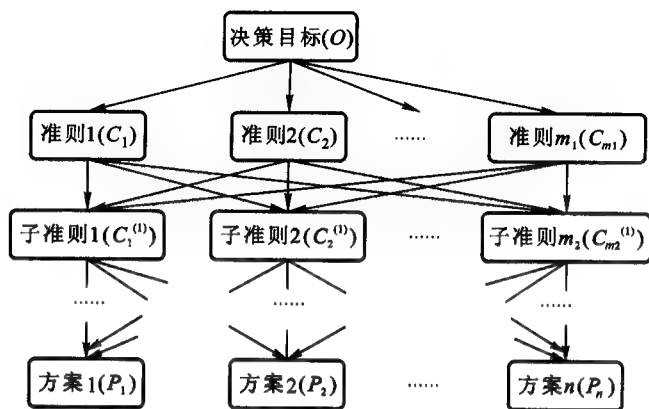


图 6-1 层次结构图

最低层为方案层(P)：方案层是为实现目标而供选择的各种措施，即为决策方案。

一般说来，各层次之间的各因素，有的相关联，有的不一定相关联；各层次的因素个数也未必一定相同。实际中，主要是根据问题的性质和各相关因素的类别来确定。

6.1.2 构造比较矩阵

构造比较矩阵主要是通过比较同一层次上的各因素对上一层相关因素的影响作用，而不是把所有因素放在一起比较，即将同一层的各因素进行两两对比。比较时采用相对尺度标准度量，尽可能地避免不同性质的因素之间相互比较的困难。同时，要尽量依据实际问题具体情况，减少由于决策人主观因素对结果造成的影响。

设要比较 n 个因素 C_1, C_2, \dots, C_n 对上一层（如目标层） O 的影响程度，即要确定它在 O 中所占的比重。对任意两个因素 C_i 和 C_j ，用 a_{ij} 表示 C_i 和 C_j 对 O 的影响程度之比，按 1—9 的比例标度来度量 a_{ij} ($i, j = 1, 2, \dots, n$)。于是，可得到两两成对比较矩阵 $A = (a_{ij})_{n \times n}$ ，又称为判断矩阵，显然

$$a_{ij} > 0, a_{ji} = \frac{1}{a_{ij}}, a_{ii} = 1 \quad (i, j = 1, 2, \dots, n),$$

因此，又称判断矩阵为正互反矩阵。

比例标度的确定： a_{ij} 取 1—9 的 9 个等级，而 a_{ji} 取 a_{ij} 的倒数（如表 6-1）。

表 6-1 比例标度值

| 标度 a_{ij} | 含 义 |
|-----------------------------------|--------------------------------------|
| 1 | C_i 与 C_j 的影响相同 |
| 3 | C_i 比 C_j 的影响稍强 |
| 5 | C_i 比 C_j 的影响强 |
| 7 | C_i 比 C_j 的影响明显地强 |
| 9 | C_i 比 C_j 的影响绝对地强 |
| 2, 4, 6, 8 | C_i 与 C_j 的影响之比在上述两个相邻等级之间 |
| $\frac{1}{2}, \dots, \frac{1}{9}$ | C_i 与 C_j 的影响之比为上面 a_{ij} 的互反数 |

由正互反矩阵的性质可知,只要确定 A 的上(或下)三角的 $\frac{n(n-1)}{2}$ 个元素即可. 在特殊情况下,如果判断矩阵 A 的元素具有传递性,即满足

$$a_{ik}a_{kj} = a_{ij} \quad (i, j, k = 1, 2, \dots, n)$$

则称 A 为一致性矩阵,简称为一致阵.

6.1.3 相对权重向量确定

1. 和法

取判断矩阵 n 个列向量归一化后的算术平均值,近似作为权重,即

$$w_i = \frac{1}{n} \sum_{j=1}^n \frac{a_{ij}}{\sum_{k=1}^n a_{kj}} \quad (i = 1, 2, \dots, n).$$

类似地,也可以对按行求和所得向量作归一化,得到相应的权重向量.

2. 求根法(几何平均法)

将 A 的各列(或行)向量求几何平均后归一化,可以近似作为权重,即

$$w_i = \frac{\left(\prod_{j=1}^n a_{ij} \right)^{\frac{1}{n}}}{\sum_{k=1}^n \left(\prod_{j=1}^n a_{kj} \right)^{\frac{1}{n}}} \quad (i = 1, 2, \dots, n).$$

3. 特征根法

设想把一大石头 Z 分成 n 个小块 c_1, c_2, \dots, c_n , 其重量分别为 w_1, w_2, \dots, w_n ,

则将 n 块小石头作两两比较,记 c_i, c_j 的相对重量为 $a_{ij} = \frac{w_i}{w_j}$ ($i, j = 1, 2, \dots, n$), 于

是可得到比较矩阵

$$A = \begin{bmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \cdots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & \frac{w_2}{w_2} & \cdots & \frac{w_2}{w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \cdots & \frac{w_n}{w_n} \end{bmatrix}.$$

显然, A 为一致性正互反矩阵, 记 $W = (w_1, w_2, \cdots, w_n)^T$, 即为权重向量. 且

$$A = W \cdot \left(\frac{1}{w_1}, \frac{1}{w_2}, \cdots, \frac{1}{w_n} \right),$$

则

$$AW = W \cdot \left(\frac{1}{w_1}, \frac{1}{w_2}, \cdots, \frac{1}{w_n} \right) W = nW$$

这表明 W 为矩阵 A 的特征向量, 且 n 为特征根.

事实上, 对于一般的判断矩阵 A 有 $A \cdot W = \lambda_{\max} W$, 这里 λ_{\max} ($\lambda_{\max} = n$) 是 A 的最大特征根, W 为 λ_{\max} 对应的特征向量.

将 W 作归一化后可近似地作为 A 的权重向量, 这种方法称为特征根法 (这是一种最常用的方法).

由代数的知识可知, 如果 A 为一致的正互反矩阵, 则有下列性质:

- (1) $\text{rank}(A) = 1$, 即 A 的每一行 (或列) 均为任一指定行 (或列) 的整数倍;
- (2) A 的最大特征根为 $\lambda_{\max} = n$, 其余的特征根均为 0;
- (3) 若 $A = (a_{ij})_{n \times n}$ 的最大特征根 $\lambda_{\max} = n$ 对应的特征向量为

$$W = (w_1, w_2, \cdots, w_n)^T,$$

则

$$a_{ij} = \frac{w_i}{w_j} \quad (i, j = 1, 2, \cdots, n).$$

由此可得下面定理

定理 6.1 设 n 阶方阵 $A > 0$, λ_{\max} 为 A 的最大特征根, 则

- (1) $\lambda_{\max} > 0$, 而且它所对应的特征向量为正向量;
- (2) λ_{\max} 为 A 的单特征根, 且 $\lambda_{\max} \geq n$;
- (3) λ_{\max} 对应的特征向量除差一个常数因子外是惟一的.

定理 6.2 n 阶正互反矩阵 $A = (a_{ij})_{n \times n}$ 是一致阵的充要条件是 $\lambda_{\max} = n$.

(略证)

6.1.4 一致性检验

通常情况下,由实际得到的判断矩阵不一定是-致的,即不一定满足传递性和一致性.实际中,也不必要求一致性绝对成立,但要求大体上是一致的,即不一致的程度应在容许的范围内.主要考查以下指标:

$$(1) \text{ 一致性指标: } CI = \frac{\lambda_{\max} - n}{n - 1}.$$

(2) 随机一致性指标: RI , 通常由实际经验给定的,如表 6-2.

表 6-2 随机一致性指标

| n | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|------|---|---|------|------|------|------|------|------|------|------|------|------|------|------|------|
| RI | 0 | 0 | 0.58 | 0.90 | 1.12 | 1.24 | 1.32 | 1.41 | 1.45 | 1.49 | 1.51 | 1.54 | 1.56 | 1.58 | 1.59 |

(3) 一致性比率指标: $CR = \frac{CI}{RI}$, 当 $CR < 0.10$ 时,认为判断矩阵的一致性是可以接受的,则 λ_{\max} 对应的特征向量可以作为排序的权重向量.此时

$$\lambda_{\max} \approx \sum_{i=1}^n \frac{(AW)_i}{nw_i} = \frac{1}{n} \sum_{i=1}^n \frac{\sum_{j=1}^n a_{ij}w_j}{w_i},$$

其中 $(AW)_i$ 表示 AW 的第 i 个分量.

6.1.5 计算组合权重和组合一致性检验

1. 组合权重向量

设第 $k-1$ 层上 n_{k-1} 个元素对总目标(最高层)的排序权重向量为

$$W^{(k-1)} = (w_1^{(k-1)}, w_2^{(k-1)}, \dots, w_{n_{k-1}}^{(k-1)})^T,$$

第 k 层上 n_k 个元素对上一层(第 $k-1$ 层)上第 j 个元素的权重向量为

$$P_j^{(k-1)} = (p_{1j}^{(k)}, p_{2j}^{(k)}, \dots, p_{n_k j}^{(k)})^T, j = 1, 2, \dots, n_{k-1},$$

则矩阵

$$P^{(k)} = [P_1^{(k)}, P_2^{(k)}, \dots, P_{n_{k-1}}^{(k)}]$$

是 $n_k \times n_{k-1}$ 阶矩阵,表示第 k 层上的元素对第 $k-1$ 层各元素的排序权向量.那么第 k 层上的元素对目标层(最高层)总排序权重向量为

$$\begin{aligned} W^{(k)} &= P^{(k)} \cdot W^{(k-1)} = [P_1^{(k)}, P_2^{(k)}, \dots, P_{n_{k-1}}^{(k)}] \cdot W^{(k-1)} \\ &= (w_1^{(k)}, w_2^{(k)}, \dots, w_{n_k}^{(k)})^T \end{aligned}$$

或

$$w_i^{(k)} = \sum_{j=1}^{n_{k-1}} p_{ij}^{(k)} w_j^{(k-1)}, i = 1, 2, \dots, n_k.$$

对任意的 $k > 2$ 有一般公式

$$W^{(k)} = P^{(k)} \cdot P^{(k-1)} \cdot \dots \cdot P^{(3)} \cdot W^{(2)} \quad (k > 2),$$

其中 $W^{(2)}$ 是第二层上各元素对目标层的总排序向量.

2. 组合一致性指标

设 k 层的一致性指标为 $CI_1^{(k)}, CI_2^{(k)}, \dots, CI_{n_{k-1}}^{(k)}$, 随机一致性指标为

$$RI_1^{(k)}, RI_2^{(k)}, \dots, RI_{n_{k-1}}^{(k)},$$

则第 k 层对目标层的(最高层)的组合一致性指标为

$$CI^{(k)} = (CI_1^{(k)}, CI_2^{(k)}, \dots, CI_{n_{k-1}}^{(k)}) \cdot W^{(k-1)}.$$

组合随机一致性指标为

$$RI^{(k)} = (RI_1^{(k)}, RI_2^{(k)}, \dots, RI_{n_{k-1}}^{(k)}) \cdot W^{(k-1)}.$$

组合一致性比率指标为

$$CR^{(k)} = CR^{(k-1)} + \frac{CI^{(k)}}{RI^{(k)}} \quad (k \geq 3).$$

当 $CR^{(k)} < 0.10$ 时,则认为整个层次的判断矩阵通过一致性检验.

6.2 一类选优排序问题^[2]

在任何一个单位(如院校、科研单位等)都有根据某些条件对所属人员进行选优的问题(如职称评定、选调职级、教学成果奖、科研成果奖等).为了使选优的结果更合理、更科学、更具有广泛的民主性,我们以某院校选优的实际问题为背景来分析研究这一问题.

6.2.1 问题的提出

设有 $N(N > 1)$ 个参评对象 P_n ($n = 1, 2, \dots, N$), 评判条件有 9 个方面, 分别记为 $C_k^{(1)}$ ($k = 1, 2, \dots, 9$), 评委会由 8 个部门组成, 分别记为 $C_i^{(2)}$ ($i = 1, 2, \dots, 8$), 其中 $C_1^{(2)}, C_2^{(2)}$ 对所有参评对象的各项条件都有评判权, 并且具有决定性作用; $C_3^{(2)}$ 对所有参评对象的条件 $C_1^{(1)}, C_2^{(1)}$ 和 $C_9^{(1)}$ 具有评判权; $C_4^{(2)}$ 对所有参评对象的条件 $C_k^{(1)}$ ($3 \leq k \leq 9$) 具有评判权; $C_5^{(2)}, C_6^{(2)}$ 和 $C_8^{(2)}$ 对部分有关的参评对象的所有条件具有评判权; $C_7^{(2)}$ 仅对有主要关系的参评对象的 $C_3^{(1)}, C_4^{(1)}$ 和 $C_9^{(1)}$ 具有评判权. 现在要解决的问题是根据各评判员对各参评对象的评判结果综合排序选优.

6.2.2 模型的假设与分析

模型假设:

(1) 各评判员按照业务主管部门的统一制定的量化标准对参评对象进行评判;

(2) 问题中所确定的评判员及权限是合理的,并具有充分的民主性;

(3) 问题中所确定的参评条件能够充分反映出参评对象的真实水平;

(4) 各评判员以参评对象的量化打分都是公平的.

模型的分析:

这是一个一般性、又有代表性的选优排序问题,鉴于这一问题所考虑的因素较多,需要在多层次多因素中相互比较,综合排序选优,我们利用层次分析法对此做出决策. 首先建立层次结构,共分四层:

最上层为目标层(O):选择优秀对象;

第二层为准则 I 层($C^{(1)}$):评优的条件,共有 9 个因素,依次记为 $C_k^{(1)}$ ($k = 1, 2, \dots, 9$);

第三层为准则 II 层($C^{(2)}$):评判员,共有 8 个因素,依次记为 $C_i^{(2)}$ ($i = 1, 2, \dots, 8$);

第四层为方案层(P): $N(N \geq 2)$ 个参评对象,依次记为 P_n ($n = 1, 2, \dots, N$).

由问题的条件可知各层次之间的关系, $C_k^{(1)}$ ($k = 1, 2, \dots, 9$) 都与 O 关联; $C_1^{(2)}, C_2^{(2)}, C_3^{(2)}, C_4^{(2)}, C_5^{(2)}, C_6^{(2)}, C_7^{(2)}, C_8^{(2)}$ 与所有 $C_k^{(1)}$ ($k = 1, 2, \dots, 9$) 关联; $C_3^{(2)}$ 与 $C_1^{(1)}, C_2^{(1)}$ 和 $C_9^{(1)}$ 关联; $C_4^{(2)}$ 与 $C_k^{(1)}$ ($3 \leq k \leq 9$) 关联; $C_7^{(2)}$ 与 $C_3^{(1)}, C_4^{(1)}, C_9^{(1)}$ 关联; 所有 $C_i^{(2)}$ ($i = 1, 2, \dots, 8$) 与全体参评对象 P_n ($n = 1, 2, \dots, N$) 关联.

6.2.3 模型的建立与求解

1. 确定准则 I 层($C^{(1)}$)对目标层(O)的权重 W_0 .

根据具体评优问题的实际,充分考虑各项条件 $C_k^{(1)}$ 在评优中所起的作用的大小,构造出成对比较矩阵 $A = (a_{ij})_{9 \times 9}$, A 是 9 阶正互反矩阵. 求 A 的最大特征值 λ_{\max} 及相应的特征向量,并对特征向量作归一化得

$$W_0 = (w_{01}, w_{02}, \dots, w_{09})^T.$$

由随机一致性指标 $RI = 1.45$, 计算一致性指标 $CI^{(1)}$ 和一致性比率指标 $CR^{(1)} = \frac{CI^{(1)}}{RI}$, 若 $CR^{(1)} < 0.1$, 则说明 W_0 可作为权向量, 否则要对 A 的元素进行调整.

2. 确定准则 II 层($C^{(2)}$)对目标层(O)的权重 W_1

(1) 求 $C^{(2)}$ 对 $C_k^{(1)}$ 的权重向量 $W_1^{(k)}$

根据各评判员对各项条件 $C_k^{(1)}$ 评判的权威性程度来确定相应的比较矩阵, 设 $C^{(2)}$ 对 $C_k^{(1)}$ 的比较矩阵为 $B_k = (b_{ij}^{(k)})$ ($k = 1, 2, \dots, 9$).

注意到如果有 $j_0 (1 \leq j_0 < 9)$ 个评判员对条件 $C_k^{(1)}$ 不具有评判权, 那么在构造比较矩阵 B_k 时先不考虑该评判员的作用, 即 B_k 的阶数应为 $9 - j_0$, 求出 B_k 的特征向量后在相应位置上补 j_0 个 0.

类似如上的方法可以求出每个 B_k 的最大特征值 $\lambda_{\max}^{(k)}$ 与相应的特征向量, 并作一致性检验, 将 B_k 的特征向量作归一化即为 $C^{(2)}$ 对 $C_k^{(1)}$ 的权重向量, 记为

$$W_1^{(k)} = (w_{11}^{(k)}, w_{12}^{(k)}, \dots, w_{18}^{(k)})^T \quad (k = 1, 2, \dots, 9). \quad (6.1)$$

若某个 B_{k_0} 不满足一致性, 则需对其进行调整, 使 $CR_{k_0}^{(2)} < 0.1 (1 \leq k_0 \leq 9)$, 且记 $CI_*^{(2)} = (CI_1^{(2)}, CI_2^{(2)}, \dots, CI_9^{(2)})$, $RI_*^{(2)} = (RI_1^{(2)}, RI_2^{(2)}, \dots, RI_9^{(2)})$.

(2) 确定 $C^{(2)}$ 层对 O 层的权重 W_1

由 (6.1) 式得

$$\overline{W}_1 = \begin{bmatrix} w_{11}^{(1)} & w_{11}^{(2)} & \dots & w_{11}^{(9)} \\ w_{12}^{(1)} & w_{12}^{(2)} & \dots & w_{12}^{(9)} \\ \vdots & \vdots & & \vdots \\ w_{18}^{(1)} & w_{18}^{(2)} & \dots & w_{18}^{(9)} \end{bmatrix}.$$

计算 $\overline{W}_1 \cdot W_0$, 并作归一化有

$$W_1 = (w_{11}, w_{12}, \dots, w_{18})^T.$$

(3) 组合一致性检验

由上可得 $CI^{(2)} = CI_*^{(2)} \cdot W_0$, $RI^{(2)} = RI_*^{(2)} \cdot W_0$, 则组合一致性比率指标为

$$CR^{(2)} = CR^{(1)} + \frac{CI^{(2)}}{RI^{(2)}}.$$

若 $CR^{(2)} < 0.1$, 则通过一致性检验, 否则调整 B_k .

3. 确定方案层 (P) 对准则 II 层 ($C^{(2)}$) 的权重 W_2

假设各评判员按统一的标准对所有参评对象的量化打分是已知的. 设各评判员对第 n 个参评对象 $P_n (1 \leq n \leq N)$ 量化打分记为矩阵 $T_n = (T_{ij}^{(n)})_{8 \times 9} (n = 1, 2, \dots, N)$, 即为参评对象 $P_n (1 \leq n \leq N)$ 的分数矩阵. 其中 $T_{ij}^{(n)}$ 表示第 i 个评判员 $C_i^{(2)}$ 对 P_n 的第 j 项条件 $C_j^{(1)}$ 的打分.

根据模型的假设, 由于每项条件 $C_k^{(1)} (k = 1, 2, \dots, 9)$ 对 O 的权重向量为 W_0 , 于是对所有参评对象 P_n 的分数矩阵作相应转化, 令

$$R_n = T_n \cdot W_0 = (r_1^{(n)}, r_2^{(n)}, \dots, r_8^{(n)})^T \quad (n = 1, 2, \dots, N),$$

即为参评对象 P_n 的水平向量, 其中 $r_k^{(n)} (1 \leq k \leq 8)$ 表示第 k 个评判员对参评对象 P_n 的综合评价指标. 记 $R'_k = (r_k^{(1)}, r_k^{(2)}, \dots, r_k^{(N)})^T (k = 1, 2, \dots, 8)$, 由此可构

造方案 P 层对 $C_k^{(2)}$ 比较矩阵 $D_k = (d_{ij}^{(k)})_{N \times N}$, 其中 $d_{ij}^{(k)} = \begin{cases} \frac{r_k^{(i)}}{r_k^{(j)}}, & \text{当 } r_k^{(j)} \neq 0 \text{ 时,} \\ 0, & \text{当 } r_k^{(j)} = 0 \text{ 时} \end{cases} (i,$

$j=1,2,\cdots,N$), 而且所有 D_k 均为一致阵. 于是可知 D_k 的最大特征值 $\lambda_{\max}^{(k)} = N$, $CR_k^{(3)} = 0$, 其任一系列向量都是 $\lambda_{\max}^{(k)}$ 的特征向量. 将其归一化可得 P 层对 $C_k^{(2)}$ 层的权重向量. 记作

$$W^{(k)} = (w_1^{(k)}, w_2^{(k)}, \cdots, w_N^{(k)})^T \quad (k=1,2,\cdots,8)$$

故 $W_2 = [W^{(1)}, W^{(2)}, \cdots, W^{(8)}]$ 为方案 P 层对准则 II 层的权重, 且一致性比率指标为 $CR^{(3)} = \sum_{k=1}^n CR_k^{(3)} = 0$.

4. 确定方案层 (P) 对目标层 (O) 的组合权重 W

由于 $C^{(2)}$ 对 O 的权重 W_1 和 P 对 $C^{(2)}$ 的权重为 W_2 , 于是 P 对 O 的权重为

$$\begin{aligned} W &= W_2 \cdot W_1 = [W^{(1)}, W^{(2)}, \cdots, W^{(8)}] \cdot W_1 \\ &= (w'_1, w'_2, \cdots, w'_N)^T. \end{aligned} \quad (6.2)$$

如果组合一致性比率指标为 $CR = CR^{(1)} + CR^{(2)} + CR^{(3)} < 0.1$, 则组合权重 W 可作为目标决策的依据.

5. 排序选优

由于 (6.2) 式中的 w'_n ($n=1,2,\cdots,N$) 是参评对象 P_n 对目标 O 层的权重, 即 w'_n 就表示参评对象 P_n 的综合水平指标, 按其大小依次排序选优.

6.2.4 模型的评价与应用

上面我们给出了一般的选优模型, 利用此模型可以解决了实际中多因素的排序选优问题, 这种决策方法较其他的方法更具有公平合理性和广泛的民主性, 可有效地排除人为因素对选优结果的影响. 为选拔优秀人才提供了科学的理论依据, 方法简便, 可操作性强, 易于实现, 应用价值高. 该模型虽然仅以由 8 个评判员, 9 项评判条件的排序选优问题提出的, 但模型可以直接推广到任何半定量和半定性多层次、多因素的排序选优问题. 模型的应用步骤如下:

(1) 由业务主管部门或领导机关制定各项条件的量化标准, 确定评判员及评判权限;

(2) 组织各评判员对所有参评对象的各项条件量化打分, 给出量化分数矩阵 $T_n = (T_{ij}^{(n)})_{8 \times 9}$ ($n=1,2,\cdots,N$);

(3) 根据各评判条件对评优目标的影响和各评判员的权限及权威性, 合理选择比较矩阵 A 和 B_k ($k=1,2,\cdots,9$) 使得更符合实际;

(4) 将分数矩阵 T_n 输入计算机;

(5) 由计算机根据该模型提供的方法进行科学计算, 最后输出排序结果;

(6) 根据排序结果选出优秀者.

6.3 合理分配住房问题^[3]

6.3.1 问题的提出

许多单位都有一套住房分配方案,一般是不同的.某院校现行住房分配方案采用“分档次加积分”的方法,其原则是:“按职级分档次,同档次的按任职时间先后排队分配住房,任职时间相同时再考虑其他条件(如工龄、爱人情况、职称、年龄大小等)适当加分,从高分到低分依次排队”.我们认为这种分配方案仍存在不合理性,例如,同档次的排队主要由任职先后确定,任职早在前,任职晚在后,即便是高职称、高学历,或夫妻双方都在同一单位(干部或职工),甚至有的为单位做出过突出贡献,但任职时间晚,则也只能排在后面.这种方案的主要问题是“按资排辈”,显然不能充分体现重视人才,鼓励先进等政策.

根据民意测验,80%以上的人认为相关条件为职级、任职时间(为任副处的时间)、工龄、职称、爱人情况、学历、年龄和奖励情况.

要解决的问题是:

请你按职级分档次,在同档次中综合考虑相关各项条件给出一种适用于任意 N 人的合理分配住房方案.用你的方案根据表 6-3 中的 40 人情况给出排队次序,并分析说明你的方案较原方案的合理性.

表 6-3 40 个人的基本情况及按原方案排序

| 人员 | 职级 | 任职时间 | 工作时间 | 职称 | 学历 | 爱人情况 | 出生年月 | 奖励加分 |
|----------|----|---------|---------|----|----|------|---------|------|
| P_1 | 8 | 1991.6 | 1971.9 | 中级 | 本科 | 院外 | 1954.9 | 0 |
| P_2 | 8 | 1992.12 | 1978.2 | 高级 | 硕士 | 院内职工 | 1957.3 | 4 |
| P_3 | 8 | 1992.12 | 1976.12 | 中级 | 硕士 | 院外 | 1955.3 | 1 |
| P_4 | 8 | 1992.12 | 1976.12 | 中级 | 大专 | 院外 | 1957.11 | 0 |
| P_5 | 8 | 1993.1 | 1974.2 | 中级 | 硕士 | 院外 | 1956.10 | 2 |
| P_6 | 8 | 1993.6 | 1973.5 | 中级 | 大专 | 院外 | 1955.10 | 0 |
| P_7 | 8 | 1993.12 | 1972.3 | 中级 | 大专 | 院内职工 | 1954.11 | 0 |
| P_8 | 8 | 1993.12 | 1977.10 | 高级 | 硕士 | 院内干部 | 1960.8 | 3 |
| P_9 | 8 | 1993.12 | 1972.12 | 中级 | 大专 | 院外 | 1954.5 | 0 |
| P_{10} | 8 | 1993.12 | 1974.8 | 高级 | 本科 | 院内职工 | 1956.3 | 4 |
| P_{11} | 8 | 1993.12 | 1974.4 | 中级 | 本科 | 院外 | 1956.12 | 0 |

续表

| 人员 | 职级 | 任职时间 | 工作时间 | 职称 | 学历 | 爱人情况 | 出生年月 | 奖励加分 |
|----------|----|----------|----------|----|-----|------|----------|------|
| P_{12} | 8 | 1993. 12 | 1975. 12 | 高级 | 硕士 | 院外 | 1958. 3 | 2 |
| P_{13} | 8 | 1993. 12 | 1975. 8 | 中级 | 大专 | 院外 | 1959. 1 | 0 |
| P_{14} | 8 | 1993. 12 | 1975. 9 | 中级 | 本科 | 院内职工 | 1956. 7 | 0 |
| P_{15} | 9 | 1994. 1 | 1978. 10 | 高级 | 本科 | 院内干部 | 1961. 11 | 5 |
| P_{16} | 9 | 1994. 6 | 1976. 11 | 高级 | 硕士 | 院内干部 | 1958. 2 | 0 |
| P_{17} | 9 | 1994. 6 | 1975. 9 | 高级 | 本科 | 院内职工 | 1959. 6 | 1 |
| P_{18} | 9 | 1994. 6 | 1975. 10 | 高级 | 本科 | 院内职工 | 1955. 11 | 6 |
| P_{19} | 9 | 1994. 6 | 1972. 12 | 初级 | 中专 | 院外 | 1956. 1 | 0 |
| P_{20} | 9 | 1994. 6 | 1974. 9 | 中级 | 大专 | 院内职工 | 1957. 1 | 0 |
| P_{21} | 9 | 1994. 6 | 1975. 2 | 高级 | 硕士 | 院外 | 1958. 11 | 2 |
| P_{22} | 8 | 1994. 6 | 1975. 9 | 中级 | 硕士 | 院内职工 | 1957. 4 | 3 |
| P_{23} | 9 | 1994. 6 | 1976. 5 | 中级 | 本科 | 院外 | 1957. 7 | 0 |
| P_{24} | 9 | 1994. 6 | 1977. 1 | 中级 | 本科 | 院内干部 | 1960. 3 | 0 |
| P_{25} | 8 | 1994. 6 | 1978. 10 | 高级 | 硕士 | 院内干部 | 1959. 5 | 2 |
| P_{26} | 9 | 1994. 6 | 1977. 5 | 中级 | 本科 | 院内职工 | 1958. 1 | 0 |
| P_{27} | 9 | 1994. 6 | 1978. 10 | 中级 | 硕士 | 院内干部 | 1963. 4 | 1 |
| P_{28} | 9 | 1994. 6 | 1978. 2 | 中级 | 本科 | 院外 | 1960. 5 | 0 |
| P_{29} | 9 | 1994. 6 | 1978. 10 | 高级 | 博士后 | 院内干部 | 1962. 4 | 5 |
| P_{30} | 9 | 1994. 6 | 1979. 9 | 中级 | 本科 | 院外 | 1962. 9 | 1 |
| P_{31} | 8 | 1994. 12 | 1975. 6 | 中级 | 大专 | 院内干部 | 1958. 7 | 0 |
| P_{32} | 8 | 1994. 12 | 1977. 10 | 高级 | 硕士 | 院内干部 | 1960. 8 | 2 |
| P_{33} | 8 | 1994. 12 | 1978. 7 | 高级 | 博士后 | 院外 | 1961. 12 | 5 |
| P_{34} | 9 | 1994. 12 | 1975. 8 | 高级 | 博士 | 院外 | 1957. 7 | 2 |
| P_{35} | 9 | 1994. 12 | 1978. 10 | 高级 | 博士 | 院内干部 | 1961. 4 | 3 |
| P_{36} | 9 | 1994. 12 | 1978. 10 | 高级 | 博士 | 院内干部 | 1962. 12 | 6 |
| P_{37} | 9 | 1994. 12 | 1978. 10 | 中级 | 本科 | 院内职工 | 1962. 12 | 0 |
| P_{38} | 9 | 1994. 12 | 1979. 10 | 中级 | 本科 | 院内干部 | 1963. 12 | 0 |
| P_{39} | 9 | 1995. 1 | 1979. 10 | 中级 | 本科 | 院内干部 | 1961. 7 | 0 |
| P_{40} | 9 | 1995. 6 | 1980. 1 | 高级 | 硕士 | 院内干部 | 1961. 3 | 4 |

6.3.2 模型的分析

由题意可知,该问题是半定量半定性、多因素的综合选优排序问题,是一个多目标决策问题,我们的主要利用层次分析法对此作出决策。

鉴于原来按任职时间先后排队的方案可能已被一部分人所接受,从某种意义上讲也有一定的合理性。现在提出要充分体现重视人才、鼓励先进等政策,但也有必要照顾到原方案合理的方面,如任职时间、工作时间、年龄的因素应重点考虑。于是,可以认为相关的8项条件在解决这一问题中所起的作用是不同的,应有轻重缓急之分。因此,假设8项条件所起的作用依次为任职时间、工作时间、职级、职称、爱人情况、学历、出生年月、奖励情况,这样能够符合大多数人的利益。任职时间早、工龄长、职级高、高职称、双职工、高学历、年龄大、受奖多的人员都能够得到充分的体现。任何一种条件的优越,在排序中都不能是绝对的优越,需要的是综合实力的优越。

由上面的分析,首先将各项条件进行量化,为了区分各条件中的档次差异,确定量化原则如下:

任职时间、工作时间、出生年月均按每月0.1分计算;职级差为1分,8级(处级)算2分,9级(副处级)算1分;职称每差一级1分,初级算1分,中级算2分,高级算3分;学历每差一档差1分,中专算1分,大专、本科、硕士、博士、博士后分别算2、3、4、5、6分;爱人情况:院外算1分,院内职工算2分,院内干部算3分;对原奖励得分再加1分。对40人的量化分数如表6-4。

表6-4 40人的量化分数表

| 人员 P_n | 任职时间 $T_n^{(1)}$ | 工作时间 $T_n^{(2)}$ | 职级 $T_n^{(3)}$ | 职称 $T_n^{(4)}$ | 爱人情况 $T_n^{(5)}$ | 学历 $T_n^{(6)}$ | 出生年月 $T_n^{(7)}$ | 奖励加分 $T_n^{(8)}$ |
|-------------|---------------------|---------------------|-------------------|-------------------|---------------------|-------------------|---------------------|---------------------|
| P_1 | 8.3 | 32.0 | 2 | 2 | 1 | 3 | 52.4 | 1 |
| P_2 | 6.5 | 24.3 | 2 | 3 | 2 | 4 | 49.4 | 5 |
| P_3 | 6.5 | 25.7 | 2 | 2 | 1 | 4 | 51.8 | 2 |
| P_4 | 6.5 | 25.7 | 2 | 2 | 1 | 2 | 48.6 | 1 |
| P_5 | 6.4 | 29.1 | 2 | 2 | 1 | 4 | 49.9 | 3 |
| P_6 | 5.9 | 30.0 | 2 | 2 | 1 | 2 | 51.1 | 1 |
| P_7 | 5.3 | 31.4 | 2 | 2 | 2 | 2 | 52.2 | 1 |
| P_8 | 5.3 | 24.7 | 2 | 3 | 3 | 4 | 45.3 | 4 |
| P_9 | 5.3 | 30.5 | 2 | 2 | 1 | 2 | 52.8 | 1 |
| P_{10} | 5.3 | 28.5 | 2 | 3 | 2 | 3 | 50.6 | 5 |

续表

| 人员 P_n | 任职时间 $T_n^{(1)}$ | 工作时间 $T_n^{(2)}$ | 职级 $T_n^{(3)}$ | 职称 $T_n^{(4)}$ | 爱人情况 $T_n^{(5)}$ | 学历 $T_n^{(6)}$ | 出生年月 $T_n^{(7)}$ | 奖励加分 $T_n^{(8)}$ |
|-------------|---------------------|---------------------|-------------------|-------------------|---------------------|-------------------|---------------------|---------------------|
| P_{11} | 5.3 | 28.9 | 2 | 2 | 1 | 3 | 49.7 | 1 |
| P_{12} | 5.3 | 26.9 | 2 | 3 | 1 | 4 | 48.2 | 3 |
| P_{13} | 5.3 | 27.3 | 2 | 2 | 1 | 2 | 47.2 | 1 |
| P_{14} | 5.3 | 27.2 | 2 | 2 | 2 | 3 | 50.2 | 1 |
| P_{15} | 5.2 | 23.5 | 1 | 3 | 3 | 3 | 43.6 | 6 |
| P_{16} | 4.7 | 25.8 | 1 | 3 | 3 | 4 | 48.3 | 1 |
| P_{17} | 4.7 | 27.2 | 1 | 3 | 2 | 3 | 46.8 | 2 |
| P_{18} | 4.7 | 27.1 | 1 | 3 | 2 | 3 | 51.0 | 7 |
| P_{19} | 4.7 | 30.5 | 1 | 1 | 1 | 1 | 50.8 | 1 |
| P_{20} | 4.7 | 28.4 | 1 | 2 | 2 | 2 | 49.6 | 1 |
| P_{21} | 4.7 | 27.7 | 1 | 3 | 1 | 4 | 47.4 | 3 |
| P_{22} | 4.7 | 27.2 | 2 | 2 | 2 | 4 | 49.3 | 4 |
| P_{23} | 4.7 | 26.4 | 1 | 2 | 1 | 3 | 49.0 | 1 |
| P_{24} | 4.7 | 25.6 | 1 | 2 | 3 | 3 | 45.8 | 1 |
| P_{25} | 4.7 | 23.5 | 2 | 3 | 3 | 4 | 46.9 | 3 |
| P_{26} | 4.7 | 26.2 | 1 | 2 | 2 | 3 | 48.4 | 1 |
| P_{27} | 4.7 | 23.5 | 1 | 2 | 3 | 4 | 42.1 | 2 |
| P_{28} | 4.7 | 24.3 | 1 | 2 | 1 | 3 | 45.6 | 1 |
| P_{29} | 4.7 | 23.5 | 1 | 3 | 3 | 6 | 43.3 | 6 |
| P_{30} | 4.7 | 22.4 | 1 | 2 | 1 | 3 | 42.8 | 2 |
| P_{31} | 4.1 | 27.5 | 2 | 2 | 3 | 2 | 47.8 | 1 |
| P_{32} | 4.1 | 24.7 | 2 | 3 | 3 | 4 | 45.3 | 3 |
| P_{33} | 4.1 | 23.8 | 2 | 3 | 1 | 6 | 43.5 | 6 |
| P_{34} | 4.1 | 27.3 | 1 | 3 | 1 | 4 | 49.0 | 3 |
| P_{35} | 4.1 | 23.5 | 1 | 3 | 3 | 5 | 44.5 | 4 |

续表

| 人员 P_n | 任职时间 $T_n^{(1)}$ | 工作时间 $T_n^{(2)}$ | 职级 $T_n^{(3)}$ | 职称 $T_n^{(4)}$ | 爱人情况 $T_n^{(5)}$ | 学历 $T_n^{(6)}$ | 出生年月 $T_n^{(7)}$ | 奖励加分 $T_n^{(8)}$ |
|-------------|---------------------|---------------------|-------------------|-------------------|---------------------|-------------------|---------------------|---------------------|
| P_{36} | 4.1 | 23.5 | 1 | 3 | 3 | 5 | 42.2 | 7 |
| P_{37} | 4.1 | 23.5 | 1 | 2 | 2 | 3 | 42.2 | 1 |
| P_{38} | 4.1 | 22.3 | 1 | 2 | 3 | 3 | 41.0 | 1 |
| P_{39} | 4.0 | 22.3 | 1 | 2 | 3 | 3 | 44.2 | 1 |
| P_{40} | 3.5 | 22.0 | 1 | 3 | 3 | 4 | 44.6 | 5 |

6.3.3 模型的假设

- (1) 题中所述的相关的 8 项条件是合理的,有关人员均无异议;
- (2) 8 项条件在分房方案中所起的作用依次为任职时间、工作时间、职级、职称、爱人情况、学历、出生年月、奖励情况;
- (3) 每个人的各项条件按统一原则均可量化,而且能够充分反映出每个人的实力;
- (4) 在量化有关任职时间、工龄、年龄时,均计算到 1998 年 5 月。

6.3.4 模型的建立

1. 建立层次结构

问题的层次结构共分三层:第一层为目标层(O):综合择优排序;第二层为准则层(C):相关条件,共有 8 个因素,依次为任职时间、工作时间、职级、职称、爱人情况、学历、出生年月、奖励情况,分别记为 C_k ($k=1,2,\dots,8$);第三层为方案层(P): N ($N \geq 2$) 个参评人员,依次记为 P_n ($n=1,2,\dots,N$)。

2. 确定准则层(C)对目标层(O)的权重 W_i

根据假设(2), C 层的 8 个因素是依次排列的,我们可以认为对决策目标的影响程度也是依次排列的,且相邻两个的影响程度之差可以认为基本相等。因此,构造比较矩阵如下:

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1/2 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1/3 & 1/2 & 1 & 2 & 3 & 4 & 5 & 6 \\ 1/4 & 1/3 & 1/2 & 1 & 2 & 3 & 4 & 5 \\ 1/5 & 1/4 & 1/3 & 1/2 & 1 & 2 & 3 & 4 \\ 1/6 & 1/5 & 1/4 & 1/3 & 1/2 & 1 & 2 & 3 \\ 1/7 & 1/6 & 1/5 & 1/4 & 1/3 & 1/2 & 1 & 2 \\ 1/8 & 1/7 & 1/6 & 1/5 & 1/4 & 1/3 & 1/2 & 1 \end{bmatrix}$$

这是一个 8 阶的正互反矩阵,经计算求得 A 的最大特征值为 $\lambda_{\max} \approx 8.28828$, 相应的特征向量作归一化有

$$W_1 = (0.331315, 0.23066, 0.157235, 0.105903, 0.0709356, 0.0476811, 0.0326976, 0.0235625)^T \quad (6.3)$$

对应的随机一致性指标 $RI_1 = 1.41$, 则一致性指标 $CI_1 = \frac{\lambda_{\max} - 8}{8 - 1} \approx 0.041183$, 一致性比率指标 $CR_1 = \frac{CI_1}{RI_1} \approx 0.029208 < 0.1$, 于是 W_1 作为 C 层对 O 层的权重向量.

3. 确定方案层(P)对准则层(C)的权重 W_2

根据问题的条件和模型的假设,对每个人各项条件的量化指标能够充分反映出每个人的综合实力.由此可以分别构造 P 层对准则 C_k 的比较矩阵

$$B_k = (b_{i,j}^{(k)})_{N \times N}, \text{ 其中 } b_{i,j}^{(k)} = \frac{T_i^{(k)}}{T_j^{(k)}} \quad (i, j = 1, 2, \dots, N; k = 1, 2, \dots, 8) \quad (6.4)$$

显然,所有的 B_k ($k = 1, 2, \dots, 8$) 均为一致阵,由一致阵的性质可知, B_k 的最大特征值 $\lambda_{\max}^{(k)} = N$, $CR_2^{(k)} = 0$, 其任一列向量都是 $\lambda_{\max}^{(k)}$ 的特征向量,将其归一化可得 P 对 C_k 的权重向量,记作

$$W^{(k)} = (w_1^{(k)}, w_2^{(k)}, \dots, w_N^{(k)})^T \quad (k = 1, 2, \dots, 8),$$

记

$$W_2 = [W^{(1)}, W^{(2)}, \dots, W^{(8)}]_{N \times 8} \quad (6.5)$$

即为 P 层对 C 层的权重,且一致性比率指标为 $CR_2 = \sum_{k=1}^8 CR_2^{(k)} = 0$.

4. 确定方案层(P)对目标层(O)的组合权重 W

由于 C 对 O 的权重 W_1 和 P 对 C 的权重 W_2 , 则 P 对 O 的权重为

$$\begin{aligned} W &= W_2 \cdot W_1 = [W^{(1)}, W^{(2)}, \dots, W^{(8)}] \cdot W_1 \\ &= (w_1, w_2, \dots, w_N)^T. \end{aligned} \quad (6.6)$$

其组合一致性比率指标为 $CR = CR_2 + CR_1 \approx 0.029208 < 0.1$, 因此,组合权重 W 可作为目标决策的依据.

5. 综合排序

由于(6.6)式中的 w_n ($n = 1, 2, \dots, N$) 是参评人员 P_n 对目标 O 层的权重, 即 w_n 就表示参评人 P_n 的综合实力指标,按其大小依次排序,就可以得到决策方案.

6.3.5 40 人的排队方案

在上面的模型中,取 $N = 40$. 40 个人的 8 项条件的量化指标如表 6-4, 由

(6.3)、(6.4)式经计算可得 P 层对 C 层的权重矩阵 W_2 , 其矩阵的每一列表示 W_2 的一列向量 $W^{(k)}$, 即 P 层对准则 C_k 的权重向量 ($k=1, 2, \dots, 8$).

由(6.5)、(6.6)和(6.3)式可得 P 对 O 组合权重为

$$\begin{aligned}
 W &= W_2 \cdot W_1 = (w_1, w_2, \dots, w_{40})^T \\
 &= (0.031\ 558\ 7, 0.030\ 078\ 2, 0.027\ 736\ 2, 0.026\ 742\ 8, 0.028\ 513\ 3, \\
 &\quad 0.026\ 733\ 2, 0.026\ 969\ 0, 0.028\ 775\ 6, 0.025\ 871\ 4, 0.028\ 666\ 8, \\
 &\quad 0.025\ 820\ 7, 0.027\ 265\ 6, 0.025\ 068\ 7, 0.026\ 363\ 6, 0.025\ 746\ 8, \\
 &\quad 0.024\ 723\ 9, 0.023\ 968\ 2, 0.025\ 151\ 4, 0.020\ 711\ 4, 0.022\ 595\ 7, \\
 &\quad 0.023\ 761\ 8, 0.026\ 382\ 1, 0.021\ 590\ 5, 0.023\ 177\ 6, 0.027\ 310\ 4, \\
 &\quad 0.022\ 445\ 4, 0.023\ 232\ 8, 0.021\ 068\ 5, 0.025\ 974\ 6, 0.020\ 827\ 5, \\
 &\quad 0.024\ 939\ 0, 0.026\ 546\ 0, 0.025\ 888\ 9, 0.022\ 699\ 7, 0.024\ 184\ 8, \\
 &\quad 0.024\ 824\ 8, 0.020\ 741\ 2, 0.021\ 365\ 1, 0.021\ 253\ 5, 0.022\ 724\ 8)^T.
 \end{aligned}$$

以 W 的 40 个分量作为 40 名参评人员的综合实力指标, 按大小依次排序, 结果如表 6-5.

表 6-5 40 人的排序结果

| | | | | | | | | | | |
|----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 人员 | P_1 | P_2 | P_3 | P_4 | P_5 | P_6 | P_7 | P_8 | P_9 | P_{10} |
| 名次 | 1 | 2 | 6 | 10 | 5 | 11 | 9 | 3 | 17 | 4 |
| 人员 | P_{11} | P_{12} | P_{13} | P_{14} | P_{15} | P_{16} | P_{17} | P_{18} | P_{19} | P_{20} |
| 名次 | 18 | 8 | 21 | 14 | 19 | 24 | 26 | 20 | 40 | 32 |
| 人员 | P_{21} | P_{22} | P_{23} | P_{24} | P_{25} | P_{26} | P_{27} | P_{28} | P_{29} | P_{30} |
| 名次 | 27 | 13 | 34 | 29 | 7 | 33 | 28 | 37 | 15 | 38 |
| 人员 | P_{31} | P_{32} | P_{33} | P_{34} | P_{35} | P_{36} | P_{37} | P_{38} | P_{39} | P_{40} |
| 名次 | 22 | 12 | 16 | 31 | 25 | 23 | 39 | 35 | 36 | 30 |

6.3.6 模型的结果分析

利用层次分析法给出了一种合理的分配方案, 用此方案综合 40 人的相关条件得到了一个排序结果. 从结果来看, 完全达到了问题的决策目标, 也使得每个人的特长和优势都得到了充分的体现. 既照顾到了任职早、工龄长、年龄大的人, 又突出了职称高、学历高、受奖多的人, 而且也考虑了双干部和双职工的利益. 但是, 每一个单项条件的优势都不是绝对的优势. 因此, 这种方案是合理的, 符合绝大多数人的利益. 譬如, P_1 在任职时间、工龄和年龄有绝对的优势, 尽管其他条件稍弱, 他仍然排在第一位. P_8 与 P_3 、 P_4 、 P_5 、 P_6 、 P_7 相比, 虽然任职时间晚, 工龄短, 年龄小, 但是, 在职称、学历、爱人情况、奖励情况都具有较强的优势, 因此, 他

排在第三位是应该的. 类似情况还有 P_{25} 、 P_{32} 、 P_{40} 等. 相反的, P_4 、 P_6 、 P_9 、 P_{19} 较其他人的任职稍早、工龄稍长、年龄稍大, 但其他条件明显的弱, 因此, 次序明显靠后也是应该的. 在多项条件相同时, 只要有一项略强, 就排在前面, 如 P_{35} 与 P_{36} , P_{38} 与 P_{39} 等. 这些都是符合决策原则的.

6.4 参考案例与参考文献

1. 参考案例

- (1) 选拔优秀队员问题——文献[1]:133—140
- (2) 足球队排名次问题——文献[4]:18—27
- (3) 电力系统的紧急修复问题——文献[5]:127—135
- (4) 会议分组问题——文献[7]:179—192

2. 参考文献

- [1] 韩中庚. 最佳组队方案及模型. 数学的实践与认识, 1997, 27(2): 133—140
- [2] 韩中庚等. 一种选优排序方案及模型. 数学的实践与认识, 2002, 32(4): 529—532
- [3] 韩中庚. 合理分配住房方案及模型. 信息工程学院学报, 1999, 18(4): 46—50
- [4] 全国大学生数学建模竞赛组委会. 全国大学生数学建模竞赛优秀论文汇编. 北京: 中国物价出版社, 2002
- [5] 叶其孝. 大学生数学建模竞赛辅导教材(一). 长沙: 湖南教育出版社, 1993
- [6] 姜启源, 谢金星等. 数学模型. 第三版. 北京: 高等教育出版社, 2003
- [7] 吴翊等. 数学建模的理论与实践. 长沙: 国防科技大学出版社, 2001
- [8] 王连芬. 层次分析法引论. 北京: 中国人民大学出版社, 1990

第七章 概率分布方法

在社会、生产、科研和生活实践中,许多问题的不确定现象都是由随机因素的影响所造成的,即将这种现象可以视为一些随机事件,而随机事件一般是按照一定的概率出现的.与此有关的随机因素的变化往往都会服从于一定的概率分布.在实际中,就是利用这些概率分布规律对问题进行研究,从而可以对所研究的实际问题做出估计、判断、预测和决策.因此,概率分布方法在解决实际问题的过程中有着非常广泛的应用.

7.1 排列与组合

7.1.1 排列

选排列:从 n 个不同的元素中,每次任取 $k(k \leq n)$ 个不同元素按次序排成一列,称为选排列,其排列种数记为 P_n^k ,即

$$P_n^k = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}.$$

全排列:从 n 个不同的元素中,每次取 n 个不同元素按次序排成一列,称为全排列,其排列种数记为 P_n^n ,即

$$P_n^n = n(n-1)(n-2)\cdots 2 \cdot 1 = n!.$$

有重复的排列:从 n 个不同的元素中,每次取 $k(k \leq n)$ 个元素,可以重复,按次序排成一列,这种排列称为有重复的排列,其排列种数为 $P_n^k = n^k(k \leq n)$.

不尽相异元素的全排列:如果在 n 个元素中,分别有 n_1, n_2, \dots, n_m 个元素相同,且 $n_1 + n_2 + \dots + n_m = n$,则这 n 个元素的全排列称为不尽相异元素的全排列,其排列种数为

$$P_n^n = \frac{n!}{n_1!n_2!\cdots n_m!}.$$

7.1.2 组合

无重复组合:从 n 个不同的元素中,每次任取 $k(k \leq n)$ 个不同元素,不考虑

其次序组合成一组,称为**组合**,其组合数记为 C_n^k ,或 $\binom{n}{k}$,即

$$C_n^k = \frac{n!}{(n-k)!k!} = \frac{P_n^k}{k!} \quad (k \leq n)$$

并且规定 $C_n^0 = 1$.

多组组合: 把 n 个不同的元素分成 m ($m \leq n$) 组,第 i 组中有 n_i ($i = 1, 2, \dots, m$) 个不同元素,且 $n_1 + n_2 + \dots + n_m = n$,这样的组合数为

$$C_n^{n_1, n_2, \dots, n_m} = \frac{n!}{n_1! n_2! \dots n_m!}.$$

有重复的组合: 从 n 个不同的元素中,每次取出 k ($k \leq n$) 个元素,可以重复,不考虑次序组合成一组,这种组合称为**有重复的组合**,其组合数为

$$C_n^{\hat{k}} = C_{n+k-1}^k \quad (k \leq n)$$

7.2 事件与概率

7.2.1 随机试验与事件

实际中,把对自然现象进行一次观察或一次科学试验统称为**试验**.如果试验可以在相同条件下重复进行多次,而且每次的试验结果是事前不可预知的,但可以知道所有可能出现的结果.则称它为一个**随机试验**,简称为**试验**.将随机试验的结果称为**随机事件**,简称为**事件**.在每次试验中必然要发生的事件称为**必然事件**,记为 Ω ,而在每次试验中不可能发生的事件称为**不可能事件**,记为 \emptyset .

如果 A, B, A_i ($i = 1, 2, \dots, n$) 都为事件,则事件有下列关系:

(1) **包含事件**: 如果 $A \subset B$,则称事件 A 包含于事件 B ,即它表示事件 A 发生必导致事件 B 发生.

(2) **相等事件**: $A = B$ 当且仅当 $A \subset B$,且 $B \subset A$.

(3) **和事件**: 事件 $A \cup B$ (或 $A + B$),表示事件 A 与事件 B 至少有一个发生.

一般地,事件 $\bigcup_{i=1}^n A_i = A_1 \cup A_2 \cup \dots \cup A_n$ 表示事件 A_1, A_2, \dots, A_n 中至少有一个发生.

(4) **差事件**: 事件 $A - B$ 表示事件 A 发生,而事件 B 不发生.

(5) **积事件**: 事件 $A \cap B$ (或 AB),表示事件 A 与 B 同时发生.一般地,事件

$\bigcap_{i=1}^n A_i = A_1 \cap A_2 \cap \dots \cap A_n$,表示事件 A_1, A_2, \dots, A_n 同时发生.

(6) **互不相容事件**: 事件 $A \cap B = \emptyset$ 表示在一次试验中事件 A 与 B 不可能

同时发生.

(7) 对立事件: 事件 $A \cap B = \emptyset$ 且 $A \cup B = \Omega$, 表示事件 A 与 B 不可能同时发生, 但又必然有其中之一发生, 记为 $B = \bar{A}$ 或 $A = \bar{B}$.

7.2.2 概率与条件概率

1. 概率

实际中, 我们在观察一个随机试验的各种事件时, 一般说来, 总会发现有些事件出现的可能性大, 有些事件出现的可能性小, 而有些事件出现的可能性彼此大致相同. 我们把刻画事件发生可能性大小的数量指标称为事件的概率. 事件 A 的概率记为 $P(A)$, 并且概率具有下列性质:

- (1) 对任一个事件 A 都有 $0 \leq P(A) \leq 1$;
- (2) $P(\Omega) = 1$;
- (3) $P(\emptyset) = 0$;
- (4) $P(\bar{A}) = 1 - P(A)$;
- (5) 若事件 A_1, A_2, \dots, A_n 互不相容, 则 $P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$.

2. 条件概率

在某些实际问题中, 除了需要知道事件 A 的概率外, 往往还需要知道在“事件 B 发生”的条件下事件 A 发生的概率, 称之为条件概率, 记为 $P(A|B)$. 若 $P(B) > 0$, 则有

$$P(A|B) = \frac{P(AB)}{P(B)}.$$

由此, 可以得到下面的结论:

(1) 若事件 A_1, A_2, \dots, A_n 互不相容, 且 $P(A_i) > 0$ ($i = 1, 2, \dots, n$), 则对任意事件 $B \subset \bigcup_{i=1}^n A_i$ 有

$$P(B) = \sum_{i=1}^n P(A_i)P(B|A_i).$$

这就是所谓的全概率公式.

(2) 若事件 A_1, A_2, \dots, A_n 互不相容, 且 $P(A_i) > 0$ ($i = 1, 2, \dots, n$), 则对任意事件 $B \subset \bigcup_{i=1}^n A_i$ 有逆概率公式:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{i=1}^n P(A_i)P(B|A_i)} \quad (i = 1, 2, \dots, n).$$

7.2.3 统计概率与几何概率

统计概率：假设在同一条件下进行 n 次试验，事件 A 发生了 m 次，则事件 A 发生的概率定义为

$$P(A) = \frac{A \text{ 发生的次数 } m}{\text{试验的总次数 } n}.$$

我们称这个概率为统计概率。

几何概率：假设区域 S 以及其中任一个可能出现的子区域 $A (A \subset S)$ 都是可以度量的，其大小分别为 $\mu(S)$ 和 $\mu(A)$ ，则事件 A 发生的概率为

$$P(A) = \frac{\mu(A)}{\mu(S)}.$$

这样计算的概率称为几何概率。

事实上，统计概率与几何概率都满足通常概率的公理和性质。

7.3 随机变量与分布函数

7.3.1 一维随机变量与分布函数

用数值表示的随机事件的函数称为随机变量。实际中任何用数值表示的随机事件都是随机变量，随机变量的函数也是随机变量。

设 ξ 为一随机变量，对任意的实数 x 有函数

$$F(x) = P(-\infty < \xi \leq x) = P(\xi \leq x)$$

称为随机变量 ξ 的分布函数。且对任意两个实数 $x_1, x_2 (x_1 < x_2)$ ，有

$$P(x_1 < \xi \leq x_2) = F(x_2) - F(x_1).$$

分布函数 $F(x)$ 具有下列性质：

- (1) $F(x)$ 是不减函数；
- (2) $0 \leq F(x) \leq 1$ ；
- (3) $F(x)$ 是右连续函数，即 $\lim_{x \rightarrow a^+} F(x) = F(a)$ 。

如果随机变量 ξ 所有取值为有限个或可列无穷个数值，则这种随机变量为离散型随机变量。非离散型的随机变量，则称为连续型随机变量。

如果 ξ 为离散型随机变量，所有的取值为 $x_k, k = 1, 2, \dots$ ，则称

$$P(\xi = x_k) = p_k, k = 1, 2, \dots$$

为随机变量 ξ 的分布列，其相应的分布函数为

$$F(x) = \sum_{x_k \leq x} p_k.$$

如果 ξ 为连续型随机变量, 则分布函数定义为

$$F(x) = \int_{-\infty}^x f(x) dx,$$

其中 $f(x)$ 为一个非负可积函数, 称之为随机变量 ξ 的分布密度, 或密度函数. 并满足下列性质:

$$(1) f(x) \geq 0;$$

$$(2) \int_{-\infty}^{+\infty} f(x) dx = 1;$$

$$(3) P(a < \xi \leq b) = F(b) - F(a) = \int_a^b f(x) dx;$$

$$(4) \text{当 } f(x) \text{ 为连续函数时有 } F'(x) = f(x).$$

7.3.2 多维随机变量与分布函数

如果 $\xi_1, \xi_2, \dots, \xi_n$ 为 n 个一维随机变量, 则称 $(\xi_1, \xi_2, \dots, \xi_n)$ 为 n 维随机变量 (或 n 维随机向量). 同样的可以分为离散型和连续型, 相应的也可以定义分布函数.

如果 $(\xi_1, \xi_2, \dots, \xi_n)$ 为连续型 n 维随机变量, 则 $(\xi_1, \xi_2, \dots, \xi_n)$ 分布函数定义为

$$F(x_1, x_2, \dots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \dots \int_{-\infty}^{x_n} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n,$$

其中 n 元函数 $f(x_1, x_2, \dots, x_n)$ 为非负可积函数, 称为 $(\xi_1, \xi_2, \dots, \xi_n)$ 的分布密度, 或 $\xi_1, \xi_2, \dots, \xi_n$ 的联合分布密度.

n 个随机变量 $\xi_1, \xi_2, \dots, \xi_n$ 为相互独立的充要条件是相应的联合分布函数可以表示为

$$F(x_1, x_2, \dots, x_n) = F_1(x_1) F_2(x_2) \dots F_n(x_n).$$

特别地, 对于常用的二维随机变量 (ξ, η) , 其分布密度函数表示为 $f(x, y)$, 分布函数为

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy.$$

两个随机变量 ξ, η 相互独立的充要条件是相应的联合分布函数可以表示为

$$F(x, y) = F_1(x) F_2(y).$$

7.3.3 随机变量的数学期望与方差

对于实际中的许多问题, 要具体求出分布函数往往是比较困难的, 然而, 很多时候并不必求出分布函数, 只需对与随机变量有关的一些数字特征进行研究. 这些数字特征虽然不能完全刻画随机变量, 但在一定程度上能够反映出随机变

量的一些重要特征. 最常用的数字特征是数学期望和方差, 其数学期望表示随机变量所有取值的平均值, 而方差表示随机变量偏离平均值的程度, 它们是随机变量的两个最重要的数字特征.

1. 数学期望

设 ξ 为离散型随机变量, 其分布列为 $P(\xi = x_k) = p_k, k = 1, 2, \dots$, 如果级数 $\sum_{k=1}^{\infty} |x_k| p_k$ 收敛, 则称 $\sum_{k=1}^{\infty} x_k p_k$ 为随机变量 ξ 的数学期望, 记为 $E\xi$, 即 $E\xi = \sum_{k=1}^{\infty} x_k p_k$.

设 ξ 为连续型随机变量, 其分布密度函数为 $f(x)$, 如果积分 $\int_{-\infty}^{+\infty} |x| f(x) dx$ 收敛, 则称 $\int_{-\infty}^{+\infty} xf(x) dx$ 为随机变量 ξ 的数学期望, 记为 $E\xi$, 即 $E\xi = \int_{-\infty}^{+\infty} xf(x) dx$.

类似地, 设 ξ 为一个随机变量, 函数 $g(\xi)$ 也是一个随机变量, 则有

(1) 若 ξ 为离散型随机变量, 且 $P(\xi = x_k) = p_k, k = 1, 2, \dots$, 如果级数 $\sum_{k=1}^{\infty} |g(x_k)| p_k$ 收敛, 则 $g(\xi)$ 的数学期望为

$$Eg(\xi) = \sum_{k=1}^{\infty} g(x_k) p_k;$$

(2) 若 ξ 为连续型随机变量, 如果积分 $\int_{-\infty}^{+\infty} |g(x)| f(x) dx$ 收敛, 则 $g(\xi)$ 的数学期望为

$$Eg(\xi) = \int_{-\infty}^{+\infty} g(x) f(x) dx.$$

对于二维的情况有类似的结果:

如果二维随机变量 (ξ, η) 分布列为 $P(\xi = x_i, \eta = y_j) = p_{ij} (i, j = 1, 2, \dots)$, 且级数 $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} |g(x_i, y_j)| p_{ij}$ 收敛, 则

$$Eg(\xi, \eta) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} g(x_i, y_j) p_{ij}.$$

若二维随机变量 (ξ, η) 的分布密度为 $f(x, y)$, 且积分 $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |g(x, y)| f(x, y) dx dy$ 收敛, 则

$$Eg(\xi, \eta) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) f(x, y) dx dy.$$

2. 方差

设 ξ 为一个随机变量, 如果 $E(\xi - E\xi)^2$ 存在, 则称其值为随机变量 ξ 的方差,

记为 $D\xi$. 显然有

$$D\xi = E(\xi - E\xi)^2 = E\xi^2 - (E\xi)^2.$$

若 ξ 为一个离散型随机变量, 且分布列为 $P(\xi = x_k) = p_k, k = 1, 2, \dots$, 则有

$$D\xi = \sum_{k=1}^{\infty} (x_k - E\xi)^2 p_k;$$

若 ξ 为一个连续型随机变量, 且分布密度为 $f(x)$, 则有

$$D\xi = \int_{-\infty}^{+\infty} (x - E\xi)^2 f(x) dx.$$

7.4 常用的概率分布及数字特征

(1) 两点分布: 设随机变量 ξ 只取 0 或 1 两个值, 它的分布列为

$$P(\xi = k) = p^k (1-p)^{1-k}, k = 0, 1,$$

则称 ξ 服从于两点分布, 且 $E\xi = p, D\xi = p(1-p)$.

(2) 二项分布: 设随机变量 ξ 可能的取值为 $0, 1, 2, \dots, n$, 且分布列为

$$P(\xi = k) = C_n^k p^k (1-p)^{1-k}, k = 0, 1, 2, \dots, n,$$

则称 ξ 服从于二项分布, 且 $E\xi = np, D\xi = np(1-p)$.

(3) 泊松 (Poisson) 分布: 设随机变量 ξ 可取所有非负整数值, 且分布列为

$$P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}, k = 0, 1, 2, \dots,$$

其中 $\lambda > 0$ 为常数, 则称 ξ 服从于泊松分布, 且 $E\xi = \lambda, D\xi = \lambda$.

(4) 均匀分布: 设 ξ 为连续型随机变量, 其分布密度为

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in [a, b], \\ 0, & x \notin [a, b], \end{cases}$$

则称 ξ 服从区间 $[a, b]$ 上的均匀分布, 且 $E\xi = \frac{a+b}{2}, D\xi = \frac{1}{12}(b-a)^2$.

(5) 正态分布: 若随机变量 ξ 分布密度函数为

$$f_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

则称 ξ 服从于正态分布 $N(\mu, \sigma^2)$, 记为 $\xi \sim N(\mu, \sigma^2)$, 且分布函数为

$$F_{\mu, \sigma}(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy,$$

其中 $E(\xi) = \mu, D(\xi) = \sigma^2$. 特别地, 当 $\mu = 0, \sigma = 1$ 时, 称其为标准的正态分布, 记为 $\xi \sim N(0, 1)$.

(6) χ^2 分布 $\chi^2(n)$: 若 n 个相互独立的随机变量 $\xi_1, \xi_2, \dots, \xi_n$ 都服从于

$N(0,1)$, 则称 $\xi = \sum_{k=1}^n \xi_k^2$ 服从于自由度为 n 的 χ^2 分布, 记为 $\xi \sim \chi^2(n)$, 其分布密度函数为

$$f_{\xi}(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & x \geq 0, \\ 0, & x < 0 \end{cases} \quad \text{且 } E(\xi) = n, D(\xi) = 2n.$$

(7) t 分布 $t(n)$: 设随机变量 $\xi \sim N(0,1)$, $\eta \sim \chi^2(n)$, 则称 $T = \xi / \sqrt{\frac{\eta}{n}}$ 服从于自由度为 n 的 t 分布, 记为 $T \sim t(n)$, 其分布密度函数为

$$f_T(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi} \Gamma\left(\frac{n}{2}\right)} \cdot \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

且 $E(T) = 0, D(T) = \frac{n}{n-2}$.

(8) F 分布: 设随机变量 $\xi \sim \chi^2(m)$, $\eta \sim \chi^2(n)$ 且相互独立, 则 $F = \frac{\xi/m}{\eta/n}$ 服从于自由度为 m 及 n 的 F 分布, 记为 $F \sim F(m, n)$, 其密度函数为

$$f_F(x) = \begin{cases} \frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)} m^{\frac{m}{2}} n^{\frac{n}{2}} \frac{x^{\frac{m}{2}-1}}{(mx+n)^{\frac{m+n}{2}}}, & x > 0, \\ 0, & x \leq 0 \end{cases}$$

且 $E(F) = \frac{n}{n-2} (n > 2), D(F) = \frac{2n^2(n+m-2)}{m(n-2)^2(n-4)} (n > 4)$.

(9) 二维正态分布: 设二维随机变量 (ξ, η) 的联合分布密度函数为

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left\{-\frac{1}{2(1-r^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2r\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\},$$

其中 $\sigma_1, \sigma_2 > 0, \mu_1, \mu_2$ 均为常数, $|r| < 1$, 则称二维随机变量 (ξ, η) 服从于二维正态分布 $N(\mu_1, \sigma_1; \mu_2, \sigma_2; r)$, 且 $E(\xi) = \mu_1, D(\xi) = \sigma_1^2, E(\eta) = \mu_2, D(\eta) = \sigma_2^2$.

7.5 足球门的危险区域问题

7.5.1 问题的提出

在足球比赛中,球员在对方球门前不同的位置起脚射门对球门的威胁是不一样的.在球门的正前方的威胁要大于在球门两侧射门;近距离的射门对球门的威胁要大于远射.已知标准球场长为 104 m,宽为 69 m;球门高为 2.44 m,宽为 7.32 m.

实际中,球员之间的基本素质可能有一定差异,但对于职业球员来讲一般可认为这种差别不大.另外,根据统计资料显示,射门时球的速度一般在 10 m/s 左右.请你结合球场和足球赛的实际情况建模分析,并研究下列问题:

(1) 针对球员在不同位置射门对球门的威胁度进行研究,并绘制出球门的危险区域;

(2) 在有一名守门员防守的情况下,对球员射门的威胁度和危险区域作进一步的研究.

7.5.2 问题的分析

根据这个问题,要确定球门的危险区域,也就是要确定球员射门最容易进球的区域.球员无论从哪个地方射门,都有进与不进两种可能,这本身就是一个随机事件,无非是哪些地方进球的可能性最大,即是最危险的区域.影响球员射门命中率的因素有很多,其中最重要的两点是球员基本素质(技术水平)和射门时的位置.对每一个球员来说,基本素质在短时间内是不可改变的,因此,我们主要是在确定条件下,对射门位置进行分析研究.也就是说,我们主要是针对同素质的球员在球场上任意一点射门时,研究其对球门的威胁程度.

某一球员在球门前某处向球门内某目标点射门时,该球员的素质和球员到目标点的距离决定了球到达目标点的概率,即命中球门的概率.事实上,当上述两个因素确定时,球飞向球门所在平面上的落点将呈现一个固定的概率分布.稍作分析容易断定,该分布应当是二维的正态分布,这是我们解决问题的关键所在.

球员从球场上某点射门时,首先必定在球门平面上确定一个目标点,射门后球依据该概率分布落入球门所在平面.将球门视为所在平面上的一个区域,在区域内对该分布进行积分,即可以得到这次射门命中球门的概率.然而,球员在选择射门的目标点时是任意的,而命中球门的概率对目标点选择有很强的依赖性.这样,我们遍历球门区域内的所有点,对命中概率作积分,将其定义为球场上某

点对球门的威胁程度,根据威胁度的大小来确定球门的危险区域.

7.5.3 模型的假设与符号说明

1. 模型的假设

- (1) 在理想状态下,认为球员是同质的,即基本素质相同,或差别不大;
- (2) 不考虑球员射门后空气、地面对球速的影响,设球速为 10 m/s ;
- (3) 球员射门只在前半场进行,为此假设前半场为有效射门区域;
- (4) 只考虑标准的球场: $(104 \times 69) \text{ m}^2$ 和球门: $(7.32 \times 2.44) \text{ m}^2$.

2. 符号说明

Ω 表示半场上的一个球门所在平面,是地面以上的半平面; D_0 表示球门内有点在平面 π 上所表示的区域,即 $D_0 \subset \Omega$; $A(x, y)$ 表示球场上的点, (x, y) 为其坐标; $B(y, z)$ 表示球门内的点, (y, z) 为其坐标; $p(y, z)$ 表示从球场上 A 点对准球门内 B 点射门时,命中球门的概率; $D(x, y)$ 表示球场上 (x, y) 点对球门的威胁度; k 表示球员的基本素质,是一个相对指标; d 表示球场上 A 点到球门内 B 点的直线距离; θ 表示直线 AB 在地面上的投影线与球门平面 π 的夹角(锐角).

7.5.4 模型的建立与求解

首先建立如图 7-1 所示的空间直角坐标系,即以球门的底边中点位置为原点 O ,地面为 xOy 面,球门所在的平面 π 为 yOz 面.

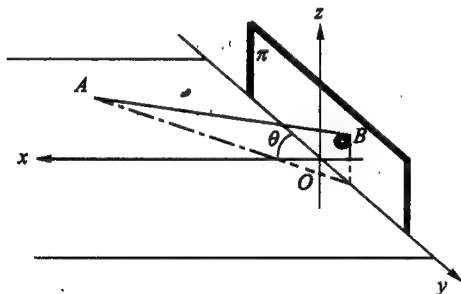


图 7-1 球门示意图

问题(1):

根据前面对问题的分析,在此假设基本素质为 k 的球员从 $A(x_0, y_0)$ 点向距离为 d 的球门内目标点 $B(y_1, z_1)$ 射门时,球在目标平面 π 上的落点呈现二维正态分布,且随机变量 y, z 是相互独立的. 其密度函数为

$$f(y, z) = \frac{1}{2\pi\sigma^2} e^{-\frac{(y-y_1)^2 + (z-z_1)^2}{2\sigma^2}}, (y, z) \in \Omega, \quad (7.1)$$

其中方差 σ 与球员素质 k 成反比, 与射门点 $A(x, y)$ 和目标点 $B(y_1, z_1)$ 之间的距离 d 成正比, 且偏角 θ 越大方差 σ 越小. 当偏角 $\theta = \frac{\pi}{2}$ 时 (即正对球门中心), 方差仅与 k, d 有关. 由此, 我们可以确定 σ 的表达式为

$$\sigma = \frac{d}{k} (\cot \theta + 1),$$

其中 $\cot \theta = \frac{|y_1 - y_0|}{x_0}$, $d = \sqrt{x_0^2 + (y_1 - y_0)^2 + z_1^2}$.

注意到, 在 (7.1) 式的密度函数中, 关于变量 y, z 是对称的, 但实际中球只能落在地面以上, 即只有 $z \geq 0$. 为了平衡这个密度函数, 我们令

$$p_D(x_0, y_0; y_1, z_1) = \iint_D f(y, z) dy dz = \iint_D \frac{1}{2\pi\sigma^2} e^{-\frac{(y-y_1)^2 + (z-z_1)^2}{2\sigma^2}} dy dz,$$

$$p_D(x_0, y_0; y_1, z_1) = \iint_D f(y, z) dy dz = \iint_D \frac{1}{2\pi\sigma^2} e^{-\frac{(y-y_1)^2 + (z-z_1)^2}{2\sigma^2}} dy dz,$$

则取两者的比值即为这次射门命中球门的概率

$$p(x_0, y_0; y_1, z_1) = \frac{p_D(x_0, y_0; y_1, z_1)}{p_D(x_0, y_0; y_1, z_1)}. \quad (7.2)$$

对命中球门的概率 (7.2) 在球门区域 D 内作积分, 定义为球场上某点 $A(x_0, y_0)$ 对球门的威胁度, 即

$$D(x_0, y_0) = \iint_D p(x_0, y_0; y_1, z_1) dy_1 dz_1.$$

综合以上分析, 对于球场上任意一点 $A(x, y)$ 关于球门的威胁度为

$$D(x, y) = \iint_D p(x, y; y_1, z_1) dy_1 dz_1,$$

其中 $p(x, y; y_1, z_1) = \frac{p_D(x, y; y_1, z_1)}{p_D(x, y; y_1, z_1)}$, $\sigma = \frac{d}{k} (\cot \theta + 1)$, $d = \sqrt{x^2 + (y_1 - y)^2 + z_1^2}$,

$$\cot \theta = \frac{|y_1 - y|}{x}.$$

要求解该问题一般是比较困难的, 只能采用数值积分的方法求解. 首先确定反应球员基本素质的参数 k , 具体的方法如下:

根据一般职业球员的情况, 我们认为一个球员在球门的正前方 ($\theta = \frac{\pi}{2}$) 距离球门 10 m 处 ($d = 10$) 向球门内的目标点劲射, 标准差应该在 1 m 以内, 即取 $\sigma = 1$, 由 $\sigma = \frac{d}{k} (\cot \theta + 1)$ 可以得到 $k = 10$. 于是, 当球员的基本素质 $k = 10$ 时, 求解该模型可以得球场上任意点对球门的威胁度, 部分特殊点的结果如表 7-1.

根据各点的威胁度的值可以做出球场上等威胁度的曲线,如图 7-2.

由图 7-2 也明显地给出了球门的危险区域.

表 7-1 球场上离散点的威胁度情况

| 位置 | (0,1) | (0,5) | (0,10) | (0,20) | (0,30) | (0,50) | (3,1) | (3,5) | (3,10) | (3,20) |
|------|----------|----------|----------|----------|----------|---------|----------|----------|----------|---------|
| 问题 1 | 14.459 6 | 14.535 1 | 12.689 1 | 8.637 1 | 5.711 5 | 2.812 1 | 11.564 9 | 13.480 1 | 11.757 8 | 7.952 7 |
| 问题 2 | 12.937 8 | 12.008 0 | 8.966 0 | 4.800 4 | 2.762 2 | 1.099 8 | 10.065 7 | 10.928 7 | 8.382 3 | 4.572 0 |
| 位置 | (3,30) | (3,50) | (5,1) | (5,5) | (5,10) | (5,20) | (5,30) | (5,50) | (10,1) | (10,5) |
| 问题 1 | 5.320 8 | 2.675 0 | 6.304 6 | 11.410 6 | 10.364 0 | 7.159 3 | 4.866 9 | 2.506 3 | 0.892 3 | 5.330 6 |
| 问题 2 | 2.663 7 | 1.076 3 | 5.897 1 | 8.946 1 | 7.234 2 | 4.125 3 | 2.453 6 | 1.013 5 | 0.815 6 | 3.919 4 |
| 位置 | (10,10) | (10,20) | (10,30) | (10,50) | (20,1) | (20,5) | (20,10) | (20,20) | (20,30) | (20,50) |
| 问题 1 | 6.465 0 | 5.242 5 | 3.817 5 | 2.118 9 | 0.060 2 | 0.876 1 | 1.847 4 | 2.426 5 | 2.188 4 | 1.481 5 |
| 问题 2 | 4.313 5 | 3.008 5 | 1.930 3 | 0.857 0 | 0.042 1 | 0.586 4 | 1.155 0 | 1.340 6 | 1.076 8 | 0.588 1 |
| 位置 | (30,1) | (30,5) | (30,10) | (30,20) | (30,30) | (30,50) | | | | |
| 问题 1 | 0.012 1 | 0.218 7 | 0.586 3 | 1.085 3 | 1.205 2 | 1.013 3 | | | | |
| 问题 2 | 0.007 0 | 0.124 3 | 0.323 3 | 0.550 6 | 0.556 4 | 0.386 7 | | | | |

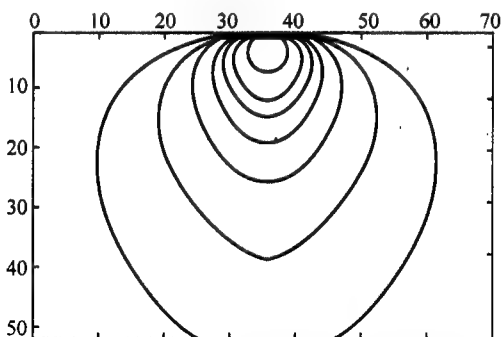


图 7-2 问题(1)的等威胁度曲线图

问题(2):

假设守门员站在射门点与两球门柱所夹角的角平分线上,即守门员站在球门在垂直射门线平面上的投影区域中心位置是最佳防守位置. 球员在球场上某点对球门内任意一点 $(y, z) \in D$ 起脚射门, 经过时间 t 到达球门平面, 球到达该点时, 守门员对球都有一个捕获的概率 $p_0(t, y, z)$, 下面先分析一下这个函数

$p_0(t, y, z)$ 的形式.

首先注意到, 当 t 一定时, $p_0(t, y, z)$ 应该是一个以守门员为中心向周围辐射衰减的二维函数, 如图 7-3(a). 图 7-3(b) 给出的是相应的等值线图.

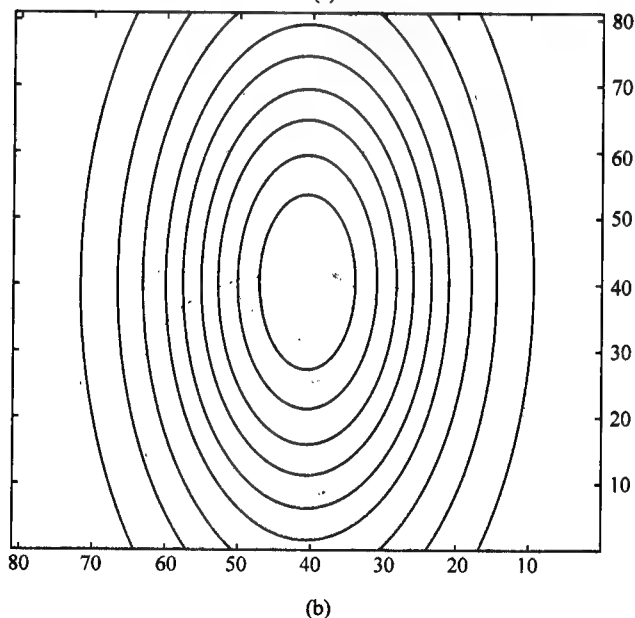
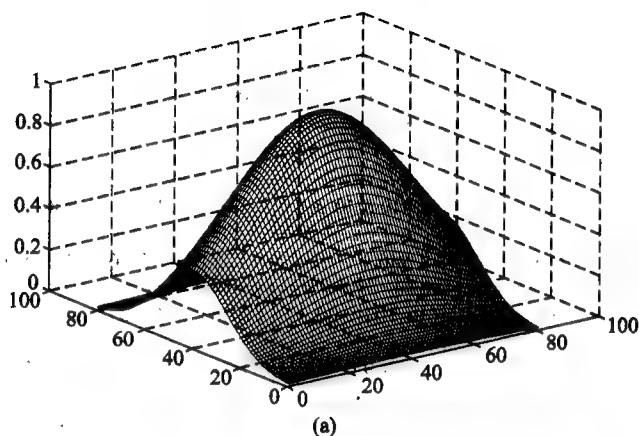


图 7-3 守门员的捕获能力示意图

当 t 变小时, 曲面的峰度应增高, 而面积减小, 如图 7-4(a) 和图 7-4(b).

由图 7-4 可以看出该曲面的形式与二维正态分布的密度函数很相似, 因此我们采用该函数形式描述这种变化趋势. 参数 t 表示从起脚射出到球到达球门

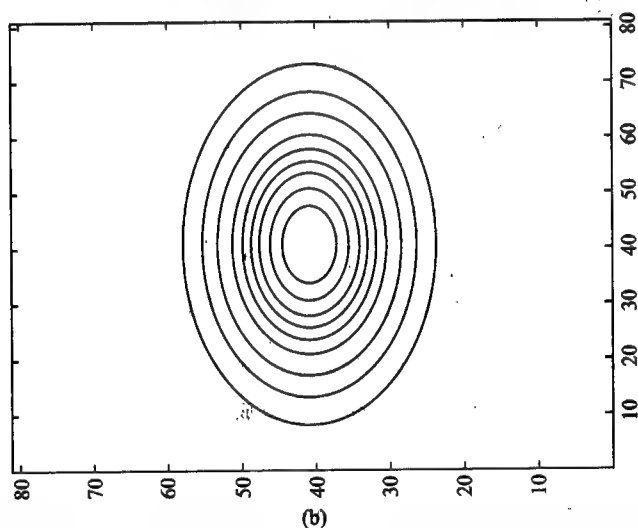
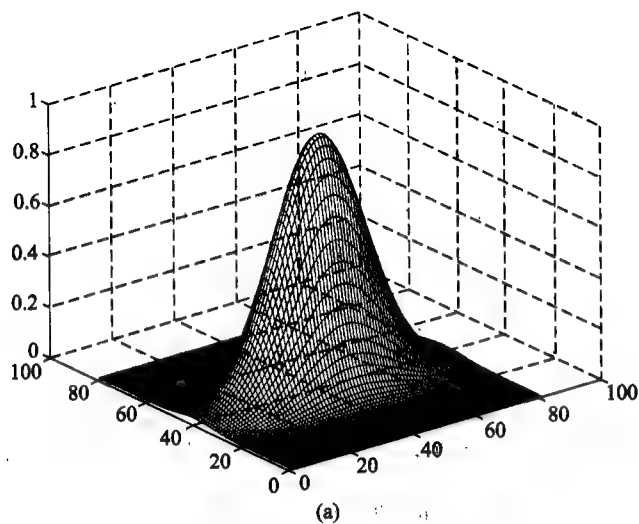


图 7-4 守门员的捕获能力随时间变化效果

的时间,也就是给守门员的反应时间,该时间越长,曲面越平滑,综上所述我们得到

$$p_0(t, y, z) = e^{-\frac{(y-a)^2 + (z-1.25)^2}{ct}},$$

其中 c 为守门员的反应系数,据专家预测,一般正常人的反应速度约为 $0.12 \sim 0.15$ s.

根据著名的“纸条实验”可得到一般人反应时间约为 $\sqrt{2}/10$ s (即设想将一张

纸条放在人的两手指之间,当纸条在重力的作用下自由下落时,由 $s = \frac{1}{2}gt^2$, 可以计算出人的反应时间). 因此,我们在此不妨取 $c = 1/7$ (实验值). 于是可得守门员防守时偏离球门中心的距离为

$$a = \frac{7.32 \sqrt{(y_0 + 3.66)^2 + x_0^2}}{\sqrt{(y_0 + 3.66)^2 + x_0^2} + \sqrt{(y_0 - 3.66)^2 + x_0^2}} - 3.66.$$

在问题(1)的基础上,对球员在球场上一点 $A(x_0, y_0)$ 射入球门的概率应修正如下:

$$\begin{aligned} p_D(x_0, y_0; y_1, z_1) &= \iint_D f(y, z) [1 - p_0(t, y, z)] dy dz \\ &= \iint_D \frac{1}{2\pi\sigma^2} e^{-\frac{(y-y_1)^2 + (z-z_1)^2}{2\sigma^2}} [1 - p_0(t, y, z)] dy dz, \end{aligned}$$

即 $p_0(t, y, z)$ 表示守门员捕获球的概率, $1 - p_0(t, y, z)$ 就表示捕不住球的概率. 于是类似地得到球场上任意一点 $A(x, y)$ 对球门的威胁度为

$$D(x, y) = \iint_D p(x, y; y_1, z_1) dy_1 dz_1,$$

其中 $p(x, y; y_1, z_1) = \frac{p_D(x, y; y_1, z_1)}{p_A(x, y; y_1, z_1)}$, $p_A(x, y; y_1, z_1)$ 同问题(1), 且 $\sigma = \frac{d}{k}(\cot \theta + 1)$,

$\cot \theta = \frac{|y_1 - y|}{x}$, $d = \sqrt{x^2 + (y_1 - y)^2 + z_1^2}$, $t = \frac{d}{v_0}$, v_0 为常数.

这里同样取进攻球员基本素质 $k = 10$, 守门员的反应系数 $c = \frac{1}{7}$, 球速 $v_0 = 10 \text{ m/s}$, 类似于问题(1)的求解可得球场上任意点对球门的威胁度, 这里给出了一些特殊点的值, 如表 7-1. 根据各点的威胁度的值也可以作出球场上等威胁度的曲线, 如图 7-5.

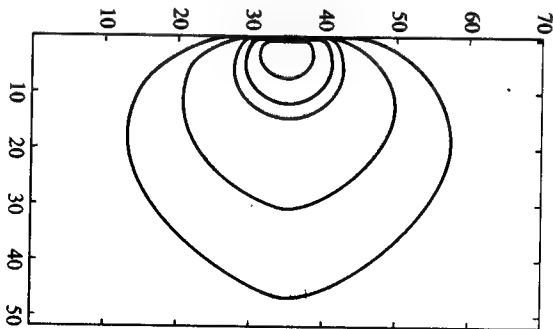


图 7-5 问题(2)的威胁度等值线

7.5.5 结果的分析与说明

比较两个问题的结果可以看出,问题(2)有防守的情况比问题(1)无防守的情况有很大的差别,问题(2)主要是守门员的作用,使得危险区明显的缩小.威胁度最大的区域还是在球门的附近,特别是正前方.由此也说明了球场上的大、小禁区设置的合理性.

本模型采用的 k 值是估算出来的,严格地讲,应该通过大量的试验按统计规律确定可能更好.我们通过计算证明了,当 k 增加(即球员的素质增强)时,对球门的威胁明显增加,危险区域变大.相反的,当 k 减小时,对球门的威胁也减小,即危险区域变小.关于防守员素质,在模型中没有考虑,是为了问题的简化.关于有多名队员的进攻和防守的情况和派兵布阵的相关问题,就更复杂了.

说明:这还是一个简化了的方法,实际中,从不同角度的位置射门,所看到的球门区域(即在垂直射门线的平面上的投影区域) D 可能不是一个矩形区域,而是一个不规则的四边形,它的形状随着射门点的变化而变化,为了简化计算还是在矩形区域 D 上作积分,这样与实际可能有些偏差.另外该问题还有多种不同解法,比如可以借助于初等几何和代数的方法,在不同的射门点进行随机模拟,通过可能射入球门的概率来定义威胁度函数,也能给出相应的结果.

7.6 参考案例与参考文献

1. 参考案例

- (1) 彩票方案的中奖率问题——文献[1]:133—140
- (2) 机票预售问题——文献[2]:144—148
- (3) 锁具的互开问题——文献[3]:58—65
- (4) 蟻的分类问题——文献[4]:37—40
- (5) 截断切割中的最优排列问题——文献[5]:120—121
- (6) 防空与空袭问题——文献[6]:195—204

2. 参考文献

- [1] 韩中庚.“彩票中的数学”问题的优化模型与评述. 工程数学学报, 2003, 20(5):107—116
- [2] 袁震东. 数学建模方法. 上海:华东师范大学出版社, 2003
- [3] 全国大学生数学建模竞赛组委会. 全国大学生数学建模竞赛优秀论文汇编. 北京:中国物价出版社, 2002
- [4] 刘承平. 数学建模方法. 北京:高等教育出版社, 2002

- [5] 赵静,但琦等. 数学建模与数学实验. 北京:高等教育出版社,2000
- [6] 《数学模型》编写组. 数学模型. 广州:华南理工大学出版社,2003
- [7] 中山大学. 概率论及数理统计. 北京:高等教育出版社,1984
- [8] 陈魁. 应用概率统计. 北京:清华大学出版社,2000

第八章 数理统计方法

在我们的现实生活中,许多问题的不确定现象都是由随机因素的影响所造成的,通常情况下,需要经过对实际中大量数据的处理或理论分析,可以确定这些随机因素所服从的概率分布,根据其概率分布规律利用一些统计方法可对所研究的问题做出估计、推断和预测.具体地讲,数理统计方法是研究从一定总体中随机抽取一部分(称为样本)的性质,来推断和预测总体的性质的一类有效方法.

8.1 统计方法的基本概念

8.1.1 总体与样本

通常我们把研究对象的全体称为**总体**,而把组成总体的每个元素称为**个体**.实际中,常常是研究对象的某个指标 X ,它一般是一个随机变量,因此,总体通常是指某个随机变量 X 取值的全体,其中每一个个体都是一个实数.如果要研究的指标不止一个,那么可以分为几个总体来研究.总体一般用大写字母 X, Y, Z 等表示.

为了要研究总体 X 的分布规律,就必须对总体进行抽样观察,根据观察结果来推断总体的性质.这种从总体 X 中抽取若干个体来观察某种数量指标的取值过程称为**抽样**.

从一个总体中,随机地抽取 n 个个体 X_1, X_2, \dots, X_n ,则称 (X_1, X_2, \dots, X_n) 为总体 X 的一个**样本**.样本中所含个体的数目 n 称为**样本的容量**.样本 (X_1, X_2, \dots, X_n) 是一个 n 维随机向量,它的每一个取值 (x_1, x_2, \dots, x_n) 称为样本的一个**观测值**,简称为**样本观测值**.

从一个总体中抽样所得样本具有下列性质:

(1) **随机性**: 对于每一次抽样,总体中的每一个个体被抽取到的机会都是相同的;

(2) **独立性**: 每一次抽样结果都是独立的.

具有这种随机性和独立性的抽样方法称为**简单随机抽样**,由此得到的样本

称为简单随机样本,我们在这里所用样本都是指简单随机样本.

实际中,统计分析的任务是由样本推断总体,具体的作法是:根据某种原则或经验假设总体服从于某种概率分布,而通过样本推断总体所服从的概率分布的若干参数.

8.1.2 统计量

样本是总体的代表,但抽取样本后,并不直接利用样本的观察值进行推断,而是把样本中所包含的有关信息集中起来进行研究,这便是针对所研究的问题构造样本的某种函数,称之为样本函数,这种样本函数称为统计量.

对于总体 X 的样本 (X_1, X_2, \dots, X_n) , 常用的统计量如下:

(1) 表示位置的统计量

$$\text{平均值: } \bar{X} = \frac{1}{n} \sum_{i=1}^n X_i.$$

中位数: 将一组数据按大小顺序排序后位于中间位置的数.

分位数: 将满足 $F(z_\alpha) = P\{X < z_\alpha\} = \alpha$ ($0 < \alpha < 1$) 的 z_α 称为分布函数 F 的 α 分位数.

(2) 表示变异程度的统计量

$$\text{标准差: } S = \left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^{\frac{1}{2}}.$$

$$\text{方差: } S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

极差: 样本中的最大值与最小值之差, 即 $S_0 = \max_{1 \leq i \leq n} (X_i) - \min_{1 \leq i \leq n} (X_i)$.

(3) 表示分布形态的统计量

偏度: $P_1 = \frac{1}{S^3} \sum_{i=1}^n (X_i - \bar{X})^3$, 它是反映数据分布的对称性的指标.

当 $P_1 > 0$ 时称为右偏态, 即数据位于均值右边的多; 当 $P_1 < 0$ 时, 称为左偏态, 即数据分布情况与上相反, 而当 $P_1 \approx 0$ 时, 则数据分布关于均值对称.

峰度: $P_2 = \frac{1}{S^4} \sum_{i=1}^n (X_i - \bar{X})^4$, 它是反映数据形态的另一个度量, 可以计算出

正态分布峰度为 3. 若 $P_2 > 3$, 则表示数据分布有沉重的尾巴, 即说明样本中含有较多的数据远离均值, 因此, 峰度是衡量偏离正态分布的指标之一.

$$k \text{ 阶原点矩: } C_k = \frac{1}{n} \sum_{i=1}^n X_i^k, k = 1, 2, \dots.$$

$$k \text{ 阶中心矩: } D_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, k = 1, 2, \dots.$$

8.2 参数估计法

所谓的参数估计就是对总体 X 的分布函数 $F(x, \theta_1, \theta_2, \dots, \theta_k)$ 由样本 (X_1, X_2, \dots, X_n) 构造一些统计量 $\hat{\theta}_i(X_1, X_2, \dots, X_n)$ ($i=1, 2, \dots, k$) 来估计 X 中的参数 (或数字特征) θ_i ($i=1, 2, \dots, k$), 这样的统计量称为估计量, 这种方法称为参数估计法. 参数估计法可以分为点估计法和区间估计法两类.

点估计: 由样本 (X_1, X_2, \dots, X_n) 构造函数 $\hat{\theta}_i(X_1, X_2, \dots, X_n)$ 作为 θ_i 的点估计量, 称统计量 $\hat{\theta}_i$ 为参数 θ_i ($i=1, 2, \dots, k$) 的点估计量.

区间估计: 由样本 (X_1, X_2, \dots, X_n) 构造两个函数 $\hat{\theta}_{i1}(X_1, X_2, \dots, X_n)$ 和 $\hat{\theta}_{i2}(X_1, X_2, \dots, X_n)$, 把区间 $(\hat{\theta}_{i1}, \hat{\theta}_{i2})$ 作为参数 θ_i ($i=1, 2, \dots, k$) 的区间估计.

8.2.1 点估计法

1. 矩估计法

假设总体 X 的分布中含有 k 个未知数 θ_i ($i=1, 2, \dots, k$). 如果它们都可以表示为原点矩的函数, 则可以用矩估计法. 即用样本 (X_1, X_2, \dots, X_n) 的 r ($1 \leq r \leq k$) 阶原点矩去估计总体 X 的相应 r 阶原点矩或原点的函数, 将 k 个参数反解出来, 从而求出各个参数的估计值. 这就是矩估计法.

例如: 设总体 $X \sim N(\mu, \sigma^2)$, 其中 μ, σ^2 是未知参数, 由样本 (X_1, X_2, \dots, X_n) 来求 μ, σ^2 的估计值.

事实上, 由于 $N(\mu, \sigma^2)$ 中含有两个未知参数, 因此, 需考虑一、二阶原点矩

$$C_1 = E(X) = \mu,$$

$$C_2 = E(X^2) = D(X) + [E(X)]^2 = \sigma^2 + \mu^2.$$

于是按矩估计法求解方程组

$$\begin{cases} \mu = \frac{1}{n} \sum_{i=1}^n X_i, \\ \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^n X_i^2, \end{cases}$$

可得 μ, σ^2 的矩估计量为

$$\begin{cases} \hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}, \\ \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = D_2. \end{cases}$$

由此可见,总体的均值 $E(X)$ 的矩估计值是样本均值 \bar{X} , 方差 $D(X)$ 的矩估计值是样本的方差.

实际上,也可以证明:无论总体 X 服从什么分布,只要 X 的均值和方差存在,此结论都是成立的.

2. 最大似然估计法

如果已知样本 (X_1, X_2, \dots, X_n) 的一组观测值为 (x_1, x_2, \dots, x_n) , 则适当的选取参数 θ_i ($i=1, 2, \dots, k$) 的值, 使样本取这组观测值的可能性最大 (即概率最大). 即构造似然函数

$$\begin{aligned} L(\theta_1, \theta_2, \dots, \theta_k) &= P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\} \\ &= \prod_{i=1}^n P\{X_i = x_i\} = \prod_{i=1}^n p(x_i, \theta_1, \theta_2, \dots, \theta_k), \end{aligned}$$

求使 $L(\theta_1, \theta_2, \dots, \theta_k)$ 达到最大的参数值, 从而可得到参数 θ_i 的估计值 $\hat{\theta}_i$ ($i=1, 2, \dots, k$). 此估计值称为最大似然估计值. 上式中的函数 p 是样本中元素的分布密度.

求最大似然估计的问题, 就是求似然函数 $L(\theta_1, \theta_2, \dots, \theta_k)$ 的最大值问题, 即可转化为求解方程组

$$\frac{\partial L}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, k$$

或

$$\frac{\partial \ln L}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, k$$

的问题.

现在的问题是如何来评价一个估计量呢?

3. 估计量的评价

一致性: 设 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 为未知参数 θ 的估计量, 若当 $n \rightarrow \infty$, $\hat{\theta}$ 依概率收敛于 θ , 则称 θ 的估计量 $\hat{\theta}$ 是一致的.

无偏性: 设 $\hat{\theta}$ 为未知参数 θ 的估计量, 若 $E(\hat{\theta}) = \theta$, 则称 $\hat{\theta}$ 为 θ 的无偏估计量.

有效性: 设 $\hat{\theta}_1, \hat{\theta}_2$ 是未知参数 θ 的两个无偏估计量, 若 $\frac{D(\hat{\theta}_1)}{D(\hat{\theta}_2)} < 1$, 则称估计量 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 有效.

8.2.2 区间估计法

参数的点估计只是求参数近似值的一种方法. 实际中, 不仅要求出参数的

近似值,而且还需要大致估计出这个值的精确程度和可信度.

设总体 X 的分布含有未知参数 θ , 若对于给定的概率(置信水平) $1 - \alpha$ ($0 < \alpha < 1$) 和样本 (X_1, X_2, \dots, X_n) , 存在两个统计量 $\hat{\theta}_1(X_1, X_2, \dots, X_n)$ 和 $\hat{\theta}_2(X_1, X_2, \dots, X_n)$ 使得

$$P\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = 1 - \alpha,$$

则称区间 $(\hat{\theta}_1, \hat{\theta}_2)$ 为参数 θ 在置信水平为 $1 - \alpha$ 下的置信区间; $\hat{\theta}_1$ 与 $\hat{\theta}_2$ 分别称为置信下限与置信上限.

1. 正态总体期望的置信区间

(1) 已知方差 $D(X) = \sigma^2$, 求 $E(X)$ 的置信区间.

设总体 $X \sim N(\mu, \sigma^2)$, (X_1, X_2, \dots, X_n) 是 X 的一个样本, 因为 $D(X) = \sigma^2$, 则

$$\xi = \frac{\bar{X} - E(X)}{\sigma/\sqrt{n}} \sim N(0, 1).$$

给定置信水平 $1 - \alpha$, 根据正态分布的对称性, 可以得到

$$P\left\{\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq E(X) \leq \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha,$$

所以 $E(X)$ 在置信水平 $1 - \alpha$ 下的置信区间为 $\left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$, 其中

$z_{1-\frac{\alpha}{2}}$ 为 $N(0, 1)$ 的分位数.

实际上, 当样本容量 n 充分大时(至少大于 50), 无论 X 服从什么分布, $\xi = \frac{\bar{X} - E(X)}{\sigma/\sqrt{n}}$ 都近似服从 $N(0, 1)$, 因此, 仍可以用 $\left[\bar{X} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$ 作为 $E(X)$ 的置信区间.

(2) 未知方差 $D(X)$, 求 $E(X)$ 的置信区间.

用 $D(X)$ 的估计值(即样本方差) $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ 代替 $D(X)$. 由于

$$\frac{\bar{X} - E(X)}{S/\sqrt{n}} \sim t(n-1),$$

根据 t 分布对称性, 则可以得到

$$P\left\{\bar{X} - t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq E(X) \leq \bar{X} + t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha,$$

所以 $E(X)$ 在置信水平 $1 - \alpha$ 下的置信区间为 $\left[\bar{X} - t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, \bar{X} + t_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right]$, 其中

$t_{1-\frac{\alpha}{2}}$ 为 $t(n-1)$ 的分位数.

2. 正态总体方差的区间估计

设总体 $X \sim N(\mu, \sigma^2)$, (X_1, X_2, \dots, X_n) 是 X 的一个样本. 由于

$$\eta = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1),$$

分别选取 χ^2 分布关于 $\frac{\sigma}{2}$ 和 $1 - \frac{\sigma}{2}$ 的分位数 $\chi_{\frac{\alpha}{2}}^2$ 和 $\chi_{1-\frac{\alpha}{2}}^2$, 则有

$$P\left\{\frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2} \leq D(X) \leq \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2}\right\} = 1 - \alpha,$$

所以 $D(X)$ 在置信水平 $1 - \alpha$ 下的置信区间为 $\left[\frac{(n-1)S^2}{\chi_{1-\frac{\alpha}{2}}^2}, \frac{(n-1)S^2}{\chi_{\frac{\alpha}{2}}^2}\right]$, 即

$$\left[\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{1-\frac{\alpha}{2}}^2}, \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{\frac{\alpha}{2}}^2}\right].$$

8.3 方差分析法

在实际中,许多问题都涉及多因素的影响,一般要先在众多的因素中选出影响大的,以进一步做更细致的试验. 用来判断一个因素的影响“是否大”的主要方法就是方差分析法. 对于所研究对象的某一项指标 X , 往往是除了受到一些因素 A_1, A_2, \dots, A_n 的影响外,还会受到随机误差的影响. 我们需要研究相关因素 A_1, A_2, \dots, A_n 对 X 影响的显著性如何? 以及随机误差对 X 的影响有多大? 要解决这些问题,方差分析就一种有效的统计方法. 方差分析可分为单因素方差分析法和多因素方差分析法,这里主要讨论单因素的情况.

假设指标 X 只受一个因素 A 的影响,或者其他因素都不变,或在一个确定的范围内变化,相应的这类问题就是单因素方差分析要解决的问题. 对于因素 A 而言,在试验时按照相关条件划分的等级或依据所处的状态划分的范围称为因素的水平.

假设因素 A 可分为 m 个水平 $h_i (i=1, 2, \dots, m)$, 在每个水平上都做了 n 次试验,其试验数据如下:

水平 $h_i: X_{i1}, X_{i2}, \dots, X_{in}$, 其样本均值为 $\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij} (i=1, 2, \dots, m)$.

实际中,由于在一种水平 h_i 下,因素 A 是确定的,则我们可以认为不同次的试验结果 $X_{i1}, X_{i2}, \dots, X_{in}$ 之间的差异与 A 无关,即只与随机误差有关. 为此,不妨假设 X_{ij} 服从于正态分布 $N(\mu_i, \sigma^2) (i=1, 2, \dots, m; j=1, 2, \dots, n)$, 其中 $\mu_i (i=1, 2, \dots, m)$ 一般不同,但是,如果因素 A 对指标 X 的影响程度不大,则可以认为 $\mu_1 = \mu_2 = \dots = \mu_m$.

对于水平 h_i 其样本方差为

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 \quad (i = 1, 2, \dots, m),$$

其中 $\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij}$, 即为期望值 μ_i ($i = 1, 2, \dots, m$) 的估计值. 于是由 m 个水平的试验结果, 我们可以得方差 σ^2 的估计值为

$$S^2 = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2.$$

由于 $\bar{X}_i = \frac{1}{n} \sum_{j=1}^n X_{ij}$ ($i = 1, 2, \dots, m$) 之间的差异反映出了 μ_i ($i = 1, 2, \dots, m$) 之间的差异, 于是可有 $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_m$ 的样本均值和方差

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m \bar{X}_i = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n X_{ij},$$

$$S_0^2 = \frac{1}{m-1} \sum_{i=1}^m (\bar{X}_i - \bar{X})^2,$$

从而方差 S_0^2 反映了各水平差异所产生的影响.

由于问题的总偏差平方和为

$$\begin{aligned} SS_T &= \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X})^2 = \sum_{i=1}^m \sum_{j=1}^n [(\bar{X}_{ij} - \bar{X}_i) + (\bar{X}_i - \bar{X})]^2 \\ &= \sum_{i=1}^m n(\bar{X}_i - \bar{X})^2 + \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2, \end{aligned}$$

其交叉项可以验证为 0. 记 $SS_A = \sum_{i=1}^m n(\bar{X}_i - \bar{X})^2$ 表示因素 A 的水平差异所产生的影响, 称为“因素 A 平方和”. $SS_E = \sum_{i=1}^m \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2$ 表示随机误差与没有考虑到的其他因素所产生的综合影响, 称为“误差平方和”. 于是有

$$SS_T = SS_A + SS_E,$$

即“总平方和”可以分解为“因素 A 平方和”与“误差平方和”两个部分.

为了要研究因素 A 对指标 X 的影响程度, 只要研究因素 A 平方和对总平方和的影响程度, 即考虑 SS_A 对 SS_T 的影响是否显著. 注意到 $SS_T = SS_A + SS_E$, 则当 SS_T 在一定的条件下, 因素 A 的影响显著, 即若 SS_A 较大时, SS_E 就较小; 相反的, 若 SS_A 较小时, SS_E 就较大. 而且 SS_A 与 SS_E 是相互独立的, 计算可得 $E(SS_A) = (m-1)\sigma^2$ 和 $E(SS_E) = m(n-1)\sigma^2$, 于是有

$$\frac{SS_A}{\sigma^2} \sim \chi^2(m-1), \quad \frac{SS_E}{\sigma^2} \sim \chi^2(m(n-1)).$$

记 SS_A 与 SS_E 的均方分别为 $MS_A = \frac{SS_A}{m-1}$ 与 $MS_E = \frac{SS_E}{m(n-1)}$, 即都是 σ^2 的无偏估计. 因此, 我们可以构造 F 统计量

$$F = \frac{MS_A}{MS_E} = \frac{\frac{1}{m-1}SS_A}{\frac{1}{m(n-1)}SS_E},$$

则 F 服从于自由度为 $(m-1, m(n-1))$ 的 F 分布 $F(m-1, m(n-1))$. 给定显著水平 α (取 0.05 或 0.01), 查 F 分布表得 $F_\alpha(m-1, m(n-1))$, 并通过样本值计算 $F(m-1, m(n-1))$, 比较二者则有

(1) 当 $F(m-1, m(n-1)) \leq F_{0.05}(m-1, m(n-1))$ 时, 则因素 A 对 X 的影响是不显著的;

(2) 当 $F_{0.05}(m-1, m(n-1)) < F(m-1, m(n-1)) \leq F_{0.01}(m-1, m(n-1))$ 时, 则因素 A 对 X 的影响是显著的;

(3) 当 $F(m-1, m(n-1)) > F_{0.01}(m-1, m(n-1))$ 时, 则因素 A 对 X 的影响是非常显著的.

上面是对于各水平 $h_i (i=1, 2, \dots, m)$ 都做相同次数的试验的条件下进行的研究, 而对于不同的试验次数有类似的结果.

假设因素 A 可分为 m 个水平 $h_i (i=1, 2, \dots, m)$, 在每个水平上分别做了 n_1, n_2, \dots, n_m 次试验, 且 $n_1 + n_2 + \dots + n_m = n$, 其试验数据为 $X_{i1}, X_{i2}, \dots, X_{in_i} (i=1, 2, \dots, m)$, 则相应的有

$$\bar{X}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} X_{ij} \quad (i=1, 2, \dots, m), \quad \bar{X} = \frac{1}{n} \sum_{i=1}^m \sum_{j=1}^{n_i} X_{ij} = \frac{1}{n} \sum_{i=1}^m n_i \bar{X}_i,$$

而且

$$SS_T = \sum_{i=1}^m \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2, \quad SS_A = \sum_{i=1}^m n_i (\bar{X}_i - \bar{X})^2, \quad SS_E = \sum_{i=1}^m \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)^2,$$

$$\frac{SS_A}{\sigma^2} \sim \chi^2(m-1), \quad \frac{SS_E}{\sigma^2} \sim \chi^2(n-m), \quad MS_A = \frac{SS_A}{m-1}, \quad MS_E = \frac{SS_E}{n-m}.$$

构造统计量为

$$F = \frac{MS_A}{MS_E} = \frac{\frac{1}{m-1}SS_A}{\frac{1}{n-m}SS_E} \sim F(m-1, n-m).$$

给定显著水平 α (取 0.05 或 0.01), 查 F 分布表得 $F_\alpha(m-1, m(n-1))$, 并通过样本值计算 $F(m-1, m(n-1))$, 比较二者则同样可以得到相应的判断结果.

8.4 相关分析方法

在许多实际问题中,经常需要研究没有因果关系,处于平等地位,而又具有一定关联的两个变量 X 与 Y 之间的密切程度的问题,这类问题就是数理统计中的相关分析方法所研究的问题. 用相关分析方法所得出的两个变量的相关关系是指统计意义下的相关,即在统计意义下,两个变量的取值存在着一定的关联关系,这并不一定说明这两个变量本身有什么直接的关系.

8.4.1 相关系数

设有两个随机变量 X 与 Y , 根据数学期望的性质, X 与 Y 相互独立, 当且仅当

$$E[(X - EX)(Y - EY)] = E(XY) - EX \cdot EY = 0,$$

即当 $E[(X - EX)(Y - EY)] \neq 0$ 时, 必有 X 与 Y 不相互独立, 则其值就反映了这两个随机变量之间的取值关系, 我们称之为随机变量 X 与 Y 的协方差(或相关矩), 记为 $\text{Cov}(X, Y)$, 即

$$\text{Cov}(X, Y) = E[(X - EX)(Y - EY)].$$

虽然协方差反映了 X 与 Y 之间的一定关系, 但它还不能从数量上完全反映出 X 与 Y 之间的密切程度, 主要是它的大小与随机变量关于它们的数学期望的偏差有关. 即当二者中有一个与数学期望的偏差很大时, 即使二者之间相关性很小, 其协方差仍然可能很大, 反之亦然. 为此, 我们将随机变量 X 与 Y 标准化, 即考虑相应的随机变量

$$X^* = \frac{X - EX}{\sqrt{DX}}, \quad Y^* = \frac{Y - EY}{\sqrt{DY}}.$$

于是, 我们把 X^* 与 Y^* 的协方差 $\text{Cov}(X^*, Y^*)$ 称为随机变量 X 与 Y 的相关系数, 记为 $r(X, Y)$ 或 r , 即

$$r = \text{Cov}(X^*, Y^*) = \frac{\text{Cov}(X, Y)}{\sqrt{DX} \sqrt{DY}}.$$

显然, 相关系数 r 是一个无量纲的量, 且 $-1 \leq r \leq 1$. 注意到 $EX^* = EY^* = 0$, $DX^* = DY^* = 1$, 因此, 相关系数能够充分地反映了随机变量 X 与 Y 之间的相关程度.

例如: 若 $Y = aX + b$ (a, b 为常数), 则 $r = 1$ ($a > 0$) 或 $r = -1$ ($a < 0$).

于是, 我们知道相关系数 r 的大小实质上只是反映了随机变量的线性相关

程度的大小. 即当 $|r|$ 的值越大, 则表明 X 与 Y 之间的线性相关程度越大, 反之亦然. 当 $|r| = 1$ 时表明 X 与 Y 之间有完全线性关系, 称之为完全线性相关; 当 $r = 0$ 时表明 X 与 Y 不存在线性关系, 称之为非线性相关, 但此时不能说明 X 与 Y 之间不存在其他非线性的函数关系.

8.4.2 相关性检验

设有随机变量 X 与 Y , 进行了 n 次随机试验, 得到观测值分别为 (X_i, Y_i) ($i = 1, 2, \dots, n$), 且 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$. 则定义

$$R = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

为随机变量 X 与 Y 对于样本 (X_i, Y_i) ($i = 1, 2, \dots, n$) 的相关系数, 称之为样本相关系数. 实际中, 常常用样本相关系数 R 作为相关系数 r 估计值.

8.5 最优评卷问题^①

8.5.1 问题的提出

在决定竞赛的优胜者时, 通常需评判大量论文, 设有 P 份论文, 由于资金有限, 参与评审的评委人数 J 和评判时间都是有限的.

理想情况是每位评委评阅每份论文, 实际中常用逐步淘汰的办法, 由于每位评委评阅一定数量的论文, 给出评价, 之后, 用某种挑选方法淘汰一部分, 再次评审剩下的论文, 重复以上过程, 当仅剩下 W 份论文时评审终止, 例如设 $P = 100$, $J = 8$, $W = 3$, 假设有某种绝对的、评委们均无异议的评判准则, 能得到 $2W$ 篇“最好”论文. 设计一个挑选方案达到如下要求:

- (1) 按此方案评审论文, 最终得到 W 篇论文, 而且使得每位评委评阅的论文数尽可能少;
- (2) 保证最终得到的 W 篇论文在 $2W$ 篇“最好”论文中;
- (3) 注意考虑“系统偏差”;
- (4) 此方案要使得参数 P, J, W 可变化.

^① 本案例选自 1996 年美国大学生数学建模竞赛 B 题, 根据信息工程大学信息工程学院彭昌勇, 彭朝阳, 刘劲松的论文(获 MCM 一等奖)整理.

8.5.2 模型的假设与符号约定

1. 模型的假设

(1) 评审论文有一个客观标准, 评委是权威且公正的, 将严格按照此标准评审论文. 在上述标准下, 对论文从 1 到 P 随机编号, 论文 k 存在一个绝对分数 $S_k (k=1, 2, \dots, P)$, 绝对分数反映的是该篇论文的客观水平, S_k 可看成是无数评委对论文 k 评分后的平均分, 并设 S_k 服从于期望为 m , 均方差为 d 的正态分布 $N(m, d)$;

(2) 评委们按照这样一种评判准则: 每个评委阅读每份论文并给出论文的分, 最后将每份论文的 J 个分数的平均值作为论文的最后得分, 并依此排序, 前 $2W$ 篇论文为最好的, 评委对此准则无异议;

(3) 评委的评分存在系统偏差, 所有评委的系统偏差服从正态分布 $N(0, D)$, 评委 j 对论文 k 的最终评分 ζ_{jk} 服从正态分布 $N(S_k + \Delta x_j, \sigma)$, Δx_j 为评委 j 的系统偏差;

(4) 不同评委的评分是相互独立的, 同一评委对不同论文的评分也是相互独立的.

2. 符号约定

$\Delta x_{\max}^{(j)}$ 表示 j 个评委评分的最大系统误差; D 表示评委的系统偏差所成分布的均方差; σ 表示评委评分的均方差; G_i 表示 J 名评委的总工作量; G_{av} 表示每个评委平均工作量 $\left(\frac{G_i}{J}\right)$; $\overline{P}_k^l = \frac{1}{l} \sum_{i=1}^l \zeta_{j_i k}$ 表示 l 个评委 j_1, j_2, \dots, j_l 的平均分, 称为 l 标准分 ($l=1, 2, \dots, J$); $\eta_1, \eta_2, \dots, \eta_n$ 表示样本; $\bar{\eta} = \frac{1}{N} \sum_{i=1}^N \eta_i$ 表示样本均值; $S^2 = \frac{1}{N} \sum_{i=1}^N (\eta_i - \bar{\eta})^2$ 表示样本方差; n_i 表示第 i 轮筛选后剩下的论文数; C_i 表示方案 4 中第 i 轮的最后删除线; d_{2W} 表示样本中第 W 名的 J 标准分; R 表示模拟中满足最后评选出的 W 名在“最好的” $2W$ 名中的条件的成功率.

8.5.3 模型的建立与求解

下面按不同的分组规则, 分别给出四种不同的逐轮淘汰的评审方案.

方案(一):

第 1 轮: 将 p 篇论文平均分成 J 组, 每位评委评审一组并判分, 将所有 p 篇论文的得分排序, 取前 n_1 名 (如 $p=100$ 时, $n_1=12$) 进入第 2 轮.

第 2 轮: 将 n_1 篇论文分成 2 组, 每组 $\frac{J}{2}$ 名评委. 通过适当分组, 尽量使得每

组的论文都是该组评委未曾看过的(可在论文上做标志达到此目的). 综合论文第 1 轮, 第 2 轮得分取平均分, 将所有 n_1 篇论文排序, 取前 $2W$ 篇入选第 3 轮.

第 3 轮: 所有评委评阅这 $2W$ 篇论文中未曾评阅的论文, 然后按 J 标准分 (即 $\overline{P}_k^J = \frac{1}{J} \sum_{j=1}^J \zeta_{jk}$) 排序, 取前 W 名为优胜者.

方案(二):

第 1 轮: 将 p 篇论文平均分为 $\frac{J}{2}$ 组, 每组 2 名评委, 评分后, 根据 2 标准分将所有 p 篇论文排序, 取前 n_1 名入选第 2 轮 (如: $P = 100$ 时, $n_1 = 12$).

第 2 轮: 将入选的 n_1 篇论文分为 2 组, 每组 $\frac{J}{2}$ 名评委, 适当分组, 尽量使得每组的论文都是该组评委未曾看过的. 综合该论文第 1 轮, 第 2 轮得分, 取平均分统一排序, 取前 n_2 (如: $P = 100$ 时, $n_2 = 6$) 篇进入第 3 轮.

第 3 轮: 所有评委评阅这 n_2 篇中未曾阅过的论文, 然后按 J 标准分排序, 取前 W 名为优胜者.

方案(三):

第 1 轮: 将 p 篇论文分成 $\frac{J}{2}$ 组, 每组按 2 标准分排序, 取每组的前 $\frac{n_1}{\left(\frac{J}{2}\right)}$ 篇,

共 n_1 篇 (如 $P = 100$ 时; $n_1 = 12$) 进入下一轮.

第 2 轮: 将该 n_1 篇论文分为 2 组, 每组 $\frac{J}{2}$ 名评委. 分组应确保每篇论文该组所有评委都未曾评阅过, 根据 $\left(\frac{J}{2+2}\right)$ 标准分排序, 每组选前 $\frac{n_2}{2}$ 篇共 n_2 篇 (如 $P = 100$ 时, $n_2 = 6$) 进入第 3 轮.

第 3 轮: 所有评委评阅该 n_2 篇论文中未曾阅过的论文, 然后按 J 标准分排序, 取前 W 名为优胜者.

方案(四):

先随机抽取 N 篇论文, 让所有 J 位评委评阅, 根据 N 篇论文的 J 标准分在 95% 的置信水平下估计出 p 篇论文的整体实力分布, 假设服从正态分布 $N(m, d)$, 然后根据此分布, 同时考虑到抽样的误差及评委的评分误差, 最后确定删除分数线 C_i .

设抽取的样本为 $\eta_1, \eta_2, \dots, \eta_N$, 其均值为 $\bar{\eta} = \frac{1}{N} \sum_{i=1}^N \eta_i$, 方差为 $S^2 =$

$\frac{1}{N} \sum_{i=1}^N (\eta_i - \bar{\eta})^2$, 根据抽样理论有

$$\sqrt{N-1} \frac{\eta - m}{S} \sim t(N-1), \frac{NS^2}{d^2} \sim \chi^2(N-1).$$

我们取 $N=30$, 可得总体实力均值 m 与均方差 d 的置信水平为 95% 的区间估计分别为 $[\bar{\eta} - 0.3S, \bar{\eta} + 0.3S]$ 和 $[0.82S, 1.35S]$.

我们将“最好”的 $2W$ 篇论文之外的论文尽可能的在最少的评价评阅后将其剔除, 为此我们首先将样本论文之外的 $P-N$ 篇论文分为 J 组, 每组的编号依次为 g_1, g_2, \dots, g_J , 将评委也依次编号, 记为 H_1, H_2, \dots, H_J .

第 1 轮: 将第 i 组论文交评委 $H_i (i=1, 2, \dots, J)$ 评阅, 其他轮中评委与论文的对对应关系如表 8-1.

表 8-1 评委评阅论文的次序

| | g_1 | g_2 | \dots | g_J |
|----------|----------|----------|----------|-----------|
| 第一轮 | H_1 | H_2 | \dots | H_J |
| 第二轮 | H_J | H_1 | \dots | H_{J-1} |
| \vdots | \vdots | \vdots | \vdots | \vdots |

评分后, 按每篇论文的得分, 即 1 标准分统一排序. 在确定删除分数线时, 应将评委评阅的最大误差考虑进去.

评委 H_j 对论文 k 的评分 ζ_{jk} 与该论文的绝对分数 S_k 的偏差由两部分组成: 一部分是评委的系统偏差 Δx_j , 另一部分是由评委评分时的随机性引起的. 不妨可设 $\zeta_{jk} \sim N(S_k + \Delta x_j, \sigma)$, 假设 J 个评委的平均分无系统误差, 即 $E\left(\frac{1}{J} \sum_{j=1}^J \zeta_{jk}\right) = S_k$, 现在只要知道 1 标准分与 J 标准分的差 $\overline{P}_k^1 - \overline{P}_k^J$ 的最大值, 再由抽样信息, 就可以确定出一个删除分数线 A_1 .

由

$$\overline{P}_k^1 - \overline{P}_k^J = \zeta_{j_1 k} - \frac{1}{J} \sum_{j=1}^J \zeta_{jk},$$

可得

$$E(\overline{P}_k^1 - \overline{P}_k^J) = S_k + \Delta x_{j_1} - S_k = \Delta x_{j_1}.$$

由假设(4), ζ_{jk} 是相互独立的, 所以

$$D(\overline{P}_k^1 - \overline{P}_k^J) = D\left[\frac{1}{8}(7\zeta_{j_1 k} - \sum_{\substack{j=1 \\ j \neq j_1}}^J \zeta_{jk})\right] = \frac{7}{8}\sigma^2,$$

于是 $\overline{P}_k^1 - \overline{P}_k^J \sim N\left(\Delta x_{j_1}, \sqrt{\frac{7}{8}}\sigma\right)$, 据 3σ 原则有

$$|\overline{P}_k^1 - \overline{P}_k^J| \leq \Delta x_{\max}^{(1)} + 3\sqrt{\frac{7}{8}}\sigma = \Delta x_{\max}^{(1)} + 2.8\sigma,$$

可取

$$\max(\overline{P}_k^1 - \overline{P}_k^J) \leq \Delta x_{\max}^{(1)} + 2.8\sigma.$$

查正态分布表,可得第 $2W$ 名的分数线为 $m + 1.89d$, m, d 只能用样本均值 $\bar{\eta}$ 和均方差 S 代替,但是同时要考虑到抽样的误差及 1 标准分与 J 标准分之差,只有充分考虑到这些误差,才能完全保证前 W 名人选第 2 轮.

综合以上得,删除分数线为

$$\begin{aligned} A_1 &= 1.89(S - 0.35S) + (\bar{\eta} - 0.3S) - \max(\overline{P}_k^1 - \overline{P}_k^J) \\ &= \bar{\eta} + 0.93S - \Delta x_{\max}^{(1)} - 2.8\sigma. \end{aligned}$$

另外,由于样本的 J 标准分已经确定,可得样本的第 $2W$ 名的 J 标准分 d_{2W} ,令

$$A_2 = d_{2W} - \max(\overline{P}_k^1 - \overline{P}_k^J),$$

则第 1 轮最后的删除分数线为 $C_1 = \max(A_1, A_2)$,删除每组中分数在 C_1 以下的论文.

第 2 轮: 将第 1 组论文中保留下来的论文送评委 H_j 评阅,第 i 组的保留论文送评委 H_{i-1} ($i=1, 2, \dots, J$) 评阅,评分后,按每篇论文在前 2 轮得分的平均值,即 2 标准分统一排序,只要找出 2 标准分与 J 标准分的最大差距,即可用类似第 1 轮的方法确定出最后的删除分数线 C_2 .事实上,由

$$\overline{P}_k^2 - \overline{P}_k^J \sim N\left(\frac{\Delta x_{j_1} + \Delta x_{j_2}}{2}, \frac{\sqrt{6}}{4}\sigma\right),$$

最后确定

$$\max(\overline{P}_k^2 - \overline{P}_k^J) = \Delta x_{\max}^{(2)} + \frac{3\sqrt{6}}{4}\sigma,$$

$$C_2 = \max(\bar{\eta} - 0.39S - \max(\overline{P}_k^2 - \overline{P}_k^J), d_{2W} - \max(\overline{P}_k^2 - \overline{P}_k^J)).$$

重复上述过程可得

$$A_1 = \bar{\eta} - 0.39S - \max(\overline{P}_k^1 - \overline{P}_k^J), A_2 = d_{2W} - \max(\overline{P}_k^1 - \overline{P}_k^J),$$

则删除分数线为 $C_t = \max(A_1, A_2)$.

注意到,按照我们的分组方法,可以确保第 t 轮后都能得到每篇论文的 t 标准分,当 $t \geq \frac{J}{2}$ 时,就使得每篇论文都有已超过半数的评委评阅,故可认为已提供了较多的信息量,因此,可将本轮的 n_{t-1} 篇论文按 t 标准分排序选优.

设第 $2W$ 名的 t 标准分为 T_{2W} ,从而确定删除分数线为

$$C_t = \max(T_{2W} - \max(\overline{P}_k^t - \overline{P}_k^J), d_{2W} - \max(\overline{P}_k^t - \overline{P}_k^J)).$$

重复此过程,直至最后剩下的论文数量很少时(论文总数的 4% 以内),让每位评委都评阅他未曾评阅过的所有入选论文,将所有剩下的论文按 J 标准分排序,取前 W 名为优胜者.

这样就给出了一个具体算法,注意到 $\max(\overline{P}_k^i - \overline{P}_k^j)$ 可用 Δx_{\max}^i 与 σ 表示,因此该算法仅涉及两个未知参数 Δx_{\max}^i 与 σ ,在对上述参数合理估计的情况下,最后用计算机模拟得出了令人满意的结果.

8.5.4 模型的模拟检验及性能分析

方案(一):

检验: 在 $P=100, J=8, W=3, D=5, \sigma=5$ 条件下,对不同的 m, d, n_1 值进行了模拟,发现 m 的变化对结果影响甚微,主要数据结果如表 8-2 ($m=50$ 时).

表 8-2 方案(一)的检验结果

| $n_1 \backslash d$ | 7.5 | 10 | 13 |
|--------------------|-------|-------|-------|
| 24 | 99.9% | 100% | 100% |
| 16 | 98.8% | 100% | 100% |
| 12 | 97.0% | 99.4% | 99.8% |
| 10 | 94.6% | 98.9% | 99.6% |

考虑一般的情形时,我们推荐 $n_1=12$,可保证 R 在 99.0% 以上. 总工作量 $G_i=166$ 篇, $G_{av}=21$ 篇.

评价: (1) 此方案工作量最小;

(2) 在 $d>7.5$ 情况下,具有相当高的成功率. 一般认为论文的总方差 $d>7.5$ 的条件是很容易满足的;

(3) 因为第 1 轮仅考虑 1 名评委的意见,故系统偏差的均方差 D 和评委评分的均方差比较大时,性能会有所下降.

方案(二):

检验: 取 $P=100, J=8, W=3$ 条件下,我们对不同的 m, d 进行模拟来初步确定 n_1, n_2 的值, m 取值范围(40, 50, 60), d 取值范围(7.5, 10, 13). 主要结果如表 8-3 ($m=50$ 时).

表 8-3 方案(二)的检验结果

| | n_1 | n_2 | G_1 | G_2 | 成功率 R |
|---------|-------|-------|-------|-------|---------|
| $d=13$ | 10 | 6 | 252 | 31.5 | 99.4% |
| $d=10$ | 12 | 6 | 260 | 32.5 | 99.0% |
| $d=7.5$ | 16 | 6 | 276 | 34.5 | 99.0% |

在 d 未知时,推荐使用 $n_1 = 16, n_2 = 6$.

评价: (1) 具有较高的精度;

(2) 比方案(三)更能适应 P 较大的情形;

(3) 总工作量偏大.

方案(三):

检验: 取 $P = 100, J = 8, W = 3, D = 5, \sigma = 5$ 时,我们对不同的 m, d, n_1, n_2 值进行了模拟, m 的取值范围 $(40, 45, 50, 60)$, d 的取值范围 $(3.5, 7.5, 10, 13)$. 结果表明 m 值的变化对 n_1, n_2 影响甚微, 而 d 值变化影响较大. 在实际运用时, 要根据不同 d 值选取 n_1, n_2 的值, 主要结果如表 8-4.

表 8-4 方案(三)的检验结果

| | n_1 | n_2 | G_1 | G_2 | 成功率 R |
|-----------|-------|-------|-------|-------|---------|
| $d = 13$ | 8 | 6 | 232 | 29 | 99.8% |
| $d = 10$ | 8 | 8 | 248 | 31 | 99.9% |
| $d = 7.5$ | 12 | 8 | 264 | 33 | 99.9% |
| $d = 5$ | 12 | 8 | 264 | 33 | 99.8% |
| $d = 3$ | 16 | 16 | 296 | 37 | 99.8% |

在 d 值未知时,建议采用 $n_1 = 16, n_2 = 6$.

评价: (1) 方案(三)简单易行;

(2) 只要 n_1, n_2 值合理,即便在 d 值很小时,也能保证相当高的成功率;

(3) 因为每轮是从每组中选出前若干名,故可大大减小系统偏差的均方差 D 的影响;

(4) 工作量 $G_{\text{总}}$ 偏大.

方案(四):

检验: 在 $P = 100, J = 8, W = 3$ 的情况下,在不同的 m, d 下,用计算机分别模拟了 500 次,得出如下结果: 平均总工作量 $W_t = 1400$, 每个评委平均工作量为 $W_{\text{平}} = \frac{W_t}{J} = 157$ 篇, 即每个评委只要评阅 $\frac{175}{500} = 35\%$ 的论文, 即可评选出符合条件的前 W 名.

评价: (1) 均值 m 的变化对总工作量影响甚微;

(2) 均方差 d 对总工作量有一定的影响, 当 d 较大时, 总工作量较小; 当 d 较小时, 总工作量较大;

(3) 经过 5 轮筛选后, 不管 m, d 取何值, 剩下论文数基本保持不变, 约为 P

的4%;

(4) 成功率为100%.

8.5.5 模型的讨论与评价

1. 关于参数 P 、 J 、 W 变化的讨论

(1) 小 P 时,适宜采用方案(一)、(二)、(三);大 P ($P > 400$) 时,优先考虑方案(四),此时 J 、 W 的变化仍不影响方案(四)的可行性;

(2) 小 W 时,方案(三)能提供足够高的精度;大 W 时,方案(二)效率显得较高;

(3) 小 J 时,可采用方案(一);大 J 时,可倾向于采用方案(二)、或方案(三).

2. 模型的优点

(1) 给出了确定“最好”的2 W 篇论文的方案,此方案简单易行,令人信服;

(2) 假设合理,充分考虑了评委评分的随机因素及评委间的“系统偏差”,从而保证了模型的通用性;

(3) 根据综合评价需要,对 P 、 J 、 W 的变化可以有针对性选取一种或两三种方案的组合,机动灵活,以达到最佳效果;

(4) 对 P 很大的情形,方案(四)是非常有效的,将极大减轻评委的工作量;

(5) 对重要参数 n_1, n_2 提供了推荐值,方便适用.

3. 模型的缺点

(1) 方案(一)、(二)、(三)中,参数 n_1, n_2 的确定主要是凭主观估计给出的,这可以用统计的方法估计 m, d 来解决;

(2) 方案(四)中参数界限的估计带有较大的主观性.

8.6 参考案例与参考文献

1. 参考案例

(1) 经济增长问题——文献[1]:9—22

(2) 自动化机床问题——文献[3]:412—440

(3) 海洋噪声的频率分析问题——文献[3]:14—27

(4) 胃癌的鉴别问题——文献[4]:47—48

(5) 学生考试成绩的分布问题——文献[5]:196—197

(6) 零件参数的设计问题——文献[2]:252—275

2. 参考文献

[1] 叶其孝. 大学生数学建模竞赛辅导教材(二). 长沙:湖南教育出版

社,1997

[2] 全国大学生数学建模竞赛组委会. 全国大学生数学建模竞赛优秀论文集. 北京:中国物价出版社,2002

[3] 汪国强. 数学建模优秀案例选编. 广州:华南理工大学出版社,1998

[4] 刘承平. 数学建模方法. 北京:高等教育出版社,2002

[5] 赵静,但琦等. 数学建模与数学实验. 北京:高等教育出版社,2000

[6] 中山大学. 概率论及数理统计. 北京:高等教育出版社,1984

[7] 陈魁. 应用概率统计. 北京:清华大学出版社,2000

第九章 回归分析方法

回归分析方法是统计分析的重要组成部分,用回归分析方法来研究建模问题是一种常用的有效方法.什么是回归分析呢?大家知道:数学分析(或高等数学)是研究连续变量之间的关系,泛函分析是研究函数集之间的关系,而回归分析是研究随机变量之间的关系.回归分析方法一般与实际联系比较密切,因为随机变量的取值是随机的,大多数是通过试验得到的,这种来自于实际中与随机变量相关的数学模型的准确度(可信度)如何,需通过进一步的统计试验来判断其模型中随机变量(回归变量)的显著性,而且,往往需要经过反复地进行检验和修改模型,直到得到最佳的结果,最后应用于实际中去.

回归分析的主要内容是:

- (1) 从一组数据出发,确定这些变量(参数)间的定量关系(回归模型);
- (2) 对模型的可信度进行统计检验;
- (3) 从有关的许多变量中,判断变量的显著性(即哪些是显著的,哪些是不显著的,显著的保留,不显著的忽略);
- (4) 应用结果是对实际问题做出的判断.

回归分析的第一步,是要建立模型,即函数关系,其自变量称为回归变量,因变量称为应变量或响应变量.如果模型中只含有一个回归变量,称为一元回归模型,否则称为多元回归模型(实际中所见的大都是线性回归模型,非线性的一般可以化为线性的来处理,例如:用泰勒(Taylor)展开法作局部线性化),为了便于理解,首先讨论一元的情况.

9.1 一元线性回归方法

9.1.1 一元线性回归模型

1. 一般形式

一元回归模型的一般形式记为

$$\eta(x) = \beta_0 + \beta_1 x,$$

并设观测值为 y , 则

$$y = \beta_0 + \beta_1 x + \varepsilon, \quad (9.1)$$

其中 β_0, β_1 是未知的待定常数, 称为回归系数; x 是回归变量, 可以是随机变量, 也可以是一般变量; ε 是随机因素对响应变量 y 所产生的影响——随机误差, 也是随机变量. 为了便于作估计和假设检验, 总是假设 $E(\varepsilon) = 0, D(\varepsilon) = \sigma^2$, 亦即 $\varepsilon \sim N(0, \sigma^2)$, 则随机变量 $y \sim N(\beta_0 + \beta_1 x, \sigma^2)$.

2. 模型的分析

假设有一组试验数据 $(x_i, y_i) (i = 1, 2, \dots, n)$, 并假设 $y_i (i = 1, 2, \dots, n)$ 是相互独立的随机变量, 则有

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, 2, \dots, n,$$

其中 ε_i 是相互独立的, 且 $\varepsilon_i \sim N(0, \sigma^2), y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2) (i = 1, 2, \dots, n)$.

若用 $\hat{\beta}_0, \hat{\beta}_1$ 分别表示 β_0, β_1 的估计值, 则称 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$ 为 y 关于 x 的一元线性回归方程. 要研究的问题是:

(1) 如何根据 $(x_i, y_i) (i = 1, 2, \dots, n)$ 来求 β_0, β_1 的估计值?

(2) 如何检验回归方程的可信度呢?

要解决第一个问题, 通常采用最小二乘估计, 第二个问题采用统计检验的方法.

9.1.2 参数 β_0, β_1 的最小二乘估计

1. 最小二乘法

用最小二乘法估计 β_0, β_1 的值, 即取 β_0, β_1 的一组估计值 $\hat{\beta}_0, \hat{\beta}_1$ 使其随机误差 ε_i 的平方和达到最小, 即使 y_i 与 $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ 的拟合最佳. 若记

$$Q(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2,$$

则

$$Q(\hat{\beta}_0, \hat{\beta}_1) = \min_{\beta_0, \beta_1} Q(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2.$$

显然 $Q(\beta_0, \beta_1) \geq 0$, 且关于 β_0, β_1 可微, 则由多元函数存在极值的必要条件得

$$\begin{cases} \left. \frac{\partial Q}{\partial \beta_0} \right|_{(\hat{\beta}_0, \hat{\beta}_1)} = 0, \\ \left. \frac{\partial Q}{\partial \beta_1} \right|_{(\hat{\beta}_0, \hat{\beta}_1)} = 0, \end{cases} \quad \text{即} \quad \begin{cases} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0, \\ \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) x_i = 0, \end{cases}$$

此方程称为正规方程组, 求解可以得到
$$\begin{cases} \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \\ \hat{\beta}_1 = l_{xy}/l_{xx}, \end{cases}$$

称 $\hat{\beta}_0, \hat{\beta}_1$ 为 β_0, β_1 的最小二乘估计, 其中

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad l_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2,$$

$$l_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - \frac{1}{n} \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right).$$

2. $\hat{\beta}_0, \hat{\beta}_1$ 的性质

$$(1) \hat{\beta}_0 \sim N\left(\beta_0, \left(\frac{1}{n} + \frac{\bar{x}^2}{l_{xx}}\right)\sigma^2\right);$$

$$(2) \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{l_{xx}}\right);$$

$$(3) \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = -\frac{\bar{x}}{l_{xx}}\sigma^2.$$

事实上, $E(\hat{\beta}_0) = \beta_0, D(\hat{\beta}_0) = \left(\frac{1}{n} + \frac{\bar{x}^2}{l_{xx}}\right)\sigma^2; E(\hat{\beta}_1) = \beta_1, D(\hat{\beta}_1) = \frac{\sigma^2}{l_{xx}}$. 由

此可知 $\hat{\beta}_0, \hat{\beta}_1$ 是 β_0, β_1 的无偏估计. 从而可以得到对固定的 x 有

$$E(\hat{y}) = E(\hat{\beta}_0 + \hat{\beta}_1 x) = E(\hat{\beta}_0) + E(\hat{\beta}_1)x = \beta_0 + \beta_1 x = E(y),$$

即 \hat{y} 是 y 的无偏估计, 且有

$$\begin{aligned} D(\hat{y}) &= D(\hat{\beta}_0 + \hat{\beta}_1 x) = D(\hat{\beta}_0) + D(\hat{\beta}_1)x^2 + 2\text{Cov}(\hat{\beta}_0, \hat{\beta}_1)x \\ &= \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{l_{xx}} \right] \sigma^2, \end{aligned}$$

故 $\hat{y} \sim N\left(\beta_0 + \beta_1 x, \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{l_{xx}} \right] \sigma^2\right)$, 即 \hat{y} 是 $\eta(x)$ 的无偏估计.

9.1.3 回归方程的显著性检验

前面是根据回归方程 $y = \beta_0 + \beta_1 x$ 求出了估计值 $\hat{\beta}_0, \hat{\beta}_1$, 从而有 $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

现在的问题是: y 与 x 之间是否确实存在这种关系? 即回归方程是否一定有意义? 即当 x 变化时, y 是否为一常数? 也就是说 β_1 是否为 0? 这就需要对回归方程作显著性检验.

实际上, 只要检验 $\beta_1 = 0$ 是否为真, 这就需要建立一个检验的统计量.

先考虑总偏差的平方和 $SS_T = \sum_{i=1}^n (y_i - \bar{y})^2$, 即表示 y_1, y_2, \dots, y_n 之间的差异, 将其分解为两个部分, 即

$$SS_T = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2$$

$$= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})$$

$$\stackrel{\text{def}}{=} SS_E + SS_R$$

事实上,由正规方程组知

$$\begin{aligned} \sum_{i=1}^n (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) &= \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(\hat{\beta}_0 + \hat{\beta}_1 x_i - \bar{y}) \\ &= \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(\hat{\beta}_0 - \bar{y}) + \hat{\beta}_1 \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i \\ &= 0. \end{aligned}$$

即回归平方和为 $SS_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$, 残差平方和(或剩余平方和)为 $SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$.

实际上, SS_R 是由回归变量 x 的变化所引起的误差, 它的大小反映了 x 的重要程度, 而 SS_E 是由随机误差和其他未加控制的因素所引起的误差. 因此, 我们现在主要考虑回归平方和 SS_R 在总的平方和 SS_T 中所占的比重, 记 $R = \frac{SS_R}{SS_T}$, 称为复相关系数, 用 R 的大小来评价模型的有效性. R 越大, 则反映了回归变量与响应变量之间的函数关系越密切, 一般 $0 \leq R \leq 1$, 但 R 要多大才认为函数关系的存在呢? 这就需要给出一个临界值, 为此引进 F 统计量.

由于每一个平方和都有一个自由度(free)(即相互独立, 且服从 $N(0, 1)$ 的随机变量的个数), 用 f 表示. 则总偏差平方和的自由度为 $f_T = \text{总观测个数} - 1 = n - 1$; 回归平方和的自由度为 $f_R = \text{回归系数个数} - 1 = 2 - 1 = 1$; 残差平方和的自由度为 $f_E = f_T - f_R = n - 2$, 于是 SS_E 的均方为 $MS_E = \frac{SS_E}{n-2}$ (平方和除以自由度).

由 $\hat{\beta}_0, \hat{\beta}_1$ 的性质可以证明: 当 $\beta_1 = 0$ 时, $E(SS_R) = \sigma^2$, $E(MS_E) = E\left(\frac{SS_E}{n-2}\right) = \sigma^2$, 即说明当 $\beta_1 = 0$ 时 MS_E 是残差的无偏差估计. 在我们的假设下 (ε_i 为独立, 同服从标准正态分布), 则回归均方 $MS_R = SS_R$ 与残差均方的比值 $F = \frac{MS_R}{MS_E} = \frac{SS_R}{SS_E/n-2}$ 是 F 统计量, 即

$$F \sim F(f_R, f_E) = F(1, n-2).$$

在 $\beta_1 = 0$ 的假设下, 给定一个模型的显著水平 α (一般取 0.01 或 0.05), 可通过查表得到 F 分布的值, 记为 $F_\alpha(1, n-2)$. 如果

$$P\{F \leq F_{\alpha}(1, n-2) \mid \beta_1 = 0\} = 1 - \alpha \geq 0.95,$$

则表明 $F > F_{\alpha}(1, n-2)$ 是小概率事件, 在一次检验中是不会发生的. 如果确实算出 $F(1, n-2) > F_{\alpha}(1, n-2)$, 则说明 $\beta_1 = 0$ 的假设不成立, 即模型中一次项 $\beta_1 x$ 是必要的, 是不可少的. 换言之, 模型对水平 α 而言是显著的, 反之是不显著的.

9.1.4 回归方程的拟合检验

通过对回归方程的显著性检验, 在显著的情况下, 即说明 x 对 y 的影响是主要的, 但不能肯定 y 与 x 的关系一定是线性的, 也可能是非线性的, 也可能还存在其他的影响因素, 为此, 就需要在同一个 x_i 下进行重复试验, 检验回归方程的拟合问题.

假设对同一个 x_i , 进行 m_i 次试验, 得到观测数据 $(x_i, y_{ij}), j = 1, 2, \dots, m_i (1 \leq i \leq n)$, 即共有 $N = \sum_{i=1}^n m_i$ 组独立观测数据, 由此来检验 $\eta(x) = \beta_0 + \beta_1 x$ 是否为真.

为了建立统计量, 考虑相应的残差平方和

$$\begin{aligned} SS_E &= \sum_{i=1}^n \sum_{j=1}^{m_i} (y_{ij} - \hat{y}_i)^2 \\ &= \sum_{i=1}^n \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2 + \sum_{i=1}^n m_i (\bar{y}_i - \hat{y}_i)^2 \quad (\text{利用正规方程组}) \\ &= SS_e + SS_{Me}, \end{aligned}$$

其中 $\bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} y_{ij}$ 为第 i 组试验数据的平均值.

$SS_e = \sum_{i=1}^n \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2$ 表示试验中的随机误差平方和, 自由度为 $f_e = N - n$.

$SS_{Me} = \sum_{i=1}^n m_i (\bar{y}_i - \hat{y}_i)^2$ 表示模型中其他影响因素所产生的误差平方和, 称为模型误差平方和(或失拟平方和), 其自由度为 $f_{Me} = n - 2$.

在回归方程为真的假设下, 则有

$$y_{ij} = \beta_0 + \beta_1 x_i + \varepsilon_{ij}, j = 1, 2, \dots, m_i; i = 1, 2, \dots, n,$$

其中 ε_{ij} 是相互独立的, 且 $\varepsilon_{ij} \sim N(0, \sigma^2)$ ($i = 1, 2, \dots, n; j = 1, 2, \dots, m$). 则

$$E(SS_{Me}) = (n-2)\sigma^2, E(SS_e) = (N-n)\sigma^2,$$

即 $E\left(\frac{SS_{Me}}{n-2}\right) = \sigma^2, E\left(\frac{SS_e}{N-n}\right) = \sigma^2$, 而 SS_{Me} 与 SS_e 是相互独立的, 由 χ^2 分布的性质可知

$$\frac{SS_{Me}}{n-2} \sim \chi^2(n-2), \frac{SS_e}{N-n} \sim \chi^2(N-n),$$

因此

$$F = \frac{MSS_{Me}}{MSS_e} = \frac{SS_{Me}/n-2}{SS_e/N-n} \sim F(f_{Me}, f_e) = F(n-2, N-n)$$

可作为检验模型拟合的统计量, 即给定一个显著水平 $\alpha(0.01 \sim 0.05)$, 对应地可查表得到 F 分布值 $F_\alpha(n-2, N-n)$.

如果计算出 $F(n-2, N-n) < F_\alpha(n-2, N-n)$, 则说明模型的拟合是好的, 即其他因素所产生的误差不明显, 是不显著的.

如果计算出 $F(n-2, N-n) > F_\alpha(n-2, N-n)$, 则说明模型的拟合是不好的, 即其他因素所产生的误差超过了试验误差, 是显著的, 需要进一步改进模型. 这两种可能: 一种是 y 不是 x 的线性关系; 另一种是回归变量的个数不够, 需要增加新的变量, 究竟属于哪一种需要找出原因做进一步的改进.

以上我们讨论了一元线性模型估计和显著性、拟合性的检验方法, 对于多元线性模型也是类似的.

9.2 多元线性回归方法

9.2.1 多元线性回归模型

多元线性回归模型的一般形式为

$$\eta(u) = \beta_1 \varphi_1(u) + \beta_2 \varphi_2(u) + \cdots + \beta_m \varphi_m(u). \quad (9.2)$$

令

$$y = \beta_1 \varphi_1(u) + \beta_2 \varphi_2(u) + \cdots + \beta_m \varphi_m(u) + \varepsilon, \quad (9.3)$$

其中 ε 为随机误差, 且服从于 $N(0, \sigma^2)$, $\varphi_i(u)$ ($i=1, 2, \dots, m$) 均为实际问题的解释变量, 是已知函数.

假设作了 n 次试验, 得到 n 组观测值为

$$\begin{bmatrix} u_1 & y_1 \\ u_2 & y_2 \\ \vdots & \vdots \\ u_n & y_n \end{bmatrix},$$

代入(9.3)中可得

$$\begin{cases} y_i = \beta_1 \varphi_1(u_i) + \beta_2 \varphi_2(u_i) + \cdots + \beta_m \varphi_m(u_i) + \varepsilon_i, \\ i = 1, 2, \cdots, n, \end{cases}$$

其中 ε_i 为第 i 次试验时的随机误差, 且相互独立同服从于 $N(0, \sigma^2)$.

该模型关于回归系数 $\beta_1, \beta_2, \cdots, \beta_m$ 是线性的, u 一般是向量.

为了方便, 引入矩阵记号:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} \varphi_1(u_1) & \varphi_2(u_1) & \cdots & \varphi_m(u_1) \\ \varphi_1(u_2) & \varphi_2(u_2) & \cdots & \varphi_m(u_2) \\ \vdots & \vdots & & \vdots \\ \varphi_1(u_n) & \varphi_2(u_n) & \cdots & \varphi_m(u_n) \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix},$$

其中 X 称为模型设计矩阵, 是常数矩阵, Y 与 ε 是随机向量, 且

$$Y \sim N_n(X \cdot \beta, \sigma^2 I), \varepsilon \sim N_n(0, \sigma^2 I) \quad (I \text{ 为 } n \text{ 阶单位阵}),$$

ε 是不可观测的随机误差向量, β 是回归系数构成的向量, 是未知待定的常数向量.

下面的问题是如何估计回归系数 β , 检验模型的显著性和拟合程度.

9.2.2 回归系数 β 的最小二乘估计

选取 β 的一个估计值, 记为 $\hat{\beta}$, 使随机误差 ε 的平方和达到最小, 即

$$\begin{aligned} \min_{\beta} \varepsilon^T \cdot \varepsilon &= \min_{\beta} (Y - X \cdot \beta)^T \cdot (Y - X \cdot \beta) \\ &= (Y - X \cdot \hat{\beta})^T \cdot (Y - X \cdot \hat{\beta}) \stackrel{\text{def}}{=} Q(\hat{\beta}), \end{aligned}$$

写成分量形式

$$Q(\beta_1, \beta_2, \cdots, \beta_m) = \sum_{i=1}^n [y_i - \beta_1 \varphi_1(u_i) - \beta_2 \varphi_2(u_i) - \cdots - \beta_m \varphi_m(u_i)]^2,$$

则

$$Q(\hat{\beta}_1, \hat{\beta}_2, \cdots, \hat{\beta}_m) = \min_{\beta_i} Q(\beta_1, \beta_2, \cdots, \beta_m).$$

注意到 $Q(\beta_1, \beta_2, \cdots, \beta_m)$ 是非负二次式, 是可微的. 由多元函数取得极值的

必要条件可得 $\frac{\partial Q}{\partial \beta_j} = 0$ ($j = 1, 2, \dots, m$), 即

$$\sum_{i=1}^n [y_i - \hat{\beta}_1 \varphi_1(\mathbf{u}_i) - \hat{\beta}_2 \varphi_2(\mathbf{u}_i) - \cdots - \hat{\beta}_m \varphi_m(\mathbf{u}_i)] \varphi_j(\mathbf{u}_i) = 0 (j = 1, 2, \dots, m),$$

整理得

$$\left\{ \begin{aligned} & \left[\sum_{i=1}^n \varphi_1^2(u_i) \right] \hat{\beta}_1 + \left[\sum_{i=1}^n \varphi_1(u_i) \varphi_2(u_i) \right] \hat{\beta}_2 + \cdots + \left[\sum_{i=1}^n \varphi_1(u_i) \varphi_m(u_i) \right] \hat{\beta}_m \\ &= \sum_{i=1}^n \varphi_1(u_i) y_i, \\ & \qquad \qquad \qquad \dots \dots \dots \dots \dots \\ & \left[\sum_{i=1}^n \varphi_1(u_i) \varphi_m(u_i) \right] \hat{\beta}_1 + \left[\sum_{i=1}^n \varphi_2(u_i) \varphi_m(u_i) \right] \hat{\beta}_2 + \cdots + \left[\sum_{i=1}^n \varphi_m^2(u_i) \right] \hat{\beta}_m \\ &= \sum_{i=1}^n \varphi_m(u_i) y_i \end{aligned} \right.$$

或

$$\mathbf{X}^T \cdot \mathbf{X} \cdot \hat{\boldsymbol{\beta}} = \mathbf{X}^T \cdot \mathbf{Y},$$

称为正规方程组, 记 $A = X^T \cdot X$ 称为系数矩阵, $B = X^T \cdot Y$ 称为常数矩阵. 如果 A^{-1} 存在, 则称其为相关矩阵. 可以证明: 对任意给定的 X, Y , 正规方程组总有解, 虽然当 X 不满秩时, 其解不唯一, 但对任意一组解 $\hat{\beta}$ 都能使残差平方和最小, 即 $Q(\hat{\beta}) = \min_{\beta} Q(\beta)$.

特别地,当 X 满秩时,即 $r(X) = r(X^T \cdot X) = m$,则正规方程组的解为 $\hat{\beta} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$,即为回归系数的估计值.

因为 $Y \sim N_n(X\beta, \sigma^2 I)$, 则 $\hat{\beta}$ 也是一个随机向量, 且期望为

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}) &= E((\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{Y}) = (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T E(\mathbf{Y}) \\ &= (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot \mathbf{X}^T \cdot \mathbf{X} \cdot \boldsymbol{\beta} = \boldsymbol{\beta}, \end{aligned}$$

同理方差为 $D(\hat{\beta}) = \sigma^2 (X^T \cdot X)^{-1}$, 即 $\hat{\beta}$ 是 β 的一个无偏估计.

将 $\hat{\beta}$ 代入模型 $\eta(u)$ 中得模型的估计: $\hat{Y} = x^T \cdot \hat{\beta}$, 它是模型 $\eta(u)$ 的无偏估计, 即 $E(\hat{Y}) = E(x^T \cdot \hat{\beta}) = x^T E(\hat{\beta}) = x^T \beta = \eta$, 其中 $x = (\varphi_1(u), \varphi_2(u), \dots, \varphi_m(u))^T$.

9.2.3 回归模型的显著性检验

主要是检验模型是否一定与解释变量有密切的关系,即是否具有(9.2)式

的形式.

假设 η 不依赖于 \mathbf{u} , 即 $\eta = \beta_0$ 为常数. 同一元的情况类似, 记试验值的均值为 $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$, 其总偏差平方和为 $SS_T(y_1, y_2, \dots, y_n \text{ 之间的误差})$, 即

$$\begin{aligned} SS_T &= \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i + \hat{y}_i - \bar{y})^2 \\ &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \stackrel{\text{def}}{=} SS_E + SS_R, \end{aligned}$$

(利用正规方程组)

其中 $SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ 为残差平方和, 反映的是随机误差和其他未加控制的因素所引起的误差, 即是误差向量 $\boldsymbol{\varepsilon}$ 的估计量 $\hat{\boldsymbol{\varepsilon}}$ 的各分量的平方和, 且

$$SS_E = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = (\mathbf{Y} - \mathbf{X} \cdot \hat{\boldsymbol{\beta}})^T \cdot (\mathbf{Y} - \mathbf{X} \cdot \hat{\boldsymbol{\beta}}) = \mathbf{Y}^T \cdot \mathbf{Y} - \mathbf{Y}^T \cdot \mathbf{X} \cdot \hat{\boldsymbol{\beta}},$$

$SS_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ 为回归平方和, 是由回归变量 \mathbf{X} 的变化引起的误差.

现在主要考虑回归平方和 SS_R , 定义复相关系数为 $R = \frac{SS_R}{SS_T}$ ($0 \leq R \leq 1$), 用 R 的大小来评价模型的有效性, R 越大, 则反映了回归变量与响应之间的关系密切, 反之亦然.

要考查 R 的大小, 需要建立一个 F 统计量, 首先求出各平方和的自由度, 总偏差平方和自由度 $f_T = n - 1$, 回归平方和的自由度为 $f_R = m - 1$, 残差平方和的自由度为 $f_E = f_T - f_R = n - m$. 于是相应的均方值为

$$MS_R = \frac{1}{m-1} SS_R, \quad MS_E = \frac{1}{n-m} SS_E,$$

可以证明: 当假设 $\eta = \beta_0$ 时, 由于 $y_i \sim N(0, \sigma^2)$, 则

$$E(MS_R) = E\left(\frac{1}{m-1} SS_R\right) = \sigma^2, \quad E(MS_E) = E\left(\frac{1}{n-m} SS_E\right) = \sigma^2.$$

这就说明 MS_E 是 σ^2 的无偏估计, 即 $\frac{SS_E}{\sigma^2} \sim \chi^2(n-m)$, $\frac{SS_R}{\sigma^2} \sim \chi^2(m-1)$, 且 SS_R 与 SS_E 相互独立, 则构造 F 统计量

$$F = \frac{MS_R}{MS_E} = \frac{SS_R/m-1}{SS_E/n-m} \sim F(f_R, f_E) = F(m-1, n-m),$$

取一个显著水平 α (0.01 或 0.05), 可查表得到 $F_\alpha(m-1, n-m)$, 计算 $F(m-1, n-m)$ 与 $F_\alpha(m-1, n-m)$ 比较:

当 $F(m-1, n-m) > F_{\alpha}(m-1, n-m)$ 时, 认为模型是显著的, 则拒绝 $\eta = \beta_0$ 成立, 即 η 与 u 存在明显的函数关系.

当 $F(m-1, n-m) < F_{\alpha}(m-1, n-m)$ 时, 认为模型是不显著的, 则 $\eta = \beta_0$ 是成立的, 即 η 与 u 不存在明显的函数关系.

9.2.4 回归模型的拟合性检验

在模型的检验显著的情况下, 需要进一步地做拟合性检验, 目的是检验模型是否一定为 (9.2) 式所给的形式, 即是否还存在其他的影响因素没有考虑到?

将回归变量 u 的 n 个观测值 u_1, u_2, \dots, u_n 按相同 (或相近) 值分成 k 组, 每组的个数记为 m_1, m_2, \dots, m_k , 显然 $n = \sum_{i=1}^k m_i$, 相应地 y_1, y_2, \dots, y_n 也可分为 k 组, 即第 i 组的观测值为 (u_i, y_{ij}) ($j = 1, 2, \dots, m_i; i = 1, 2, \dots, k$). 记 $T_i = \sum_{j=1}^{m_i} y_{ij}$, 则第 i 组的平均值为 $\bar{y}_i = \frac{T_i}{m_i}$, 根据正规方程组第 i 组的试验随机误差的平方和为

$$SS_e = \sum_{i=1}^k \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2 = \sum_{i=1}^k \sum_{j=1}^{m_i} y_{ij}^2 - \sum_{i=1}^k \frac{T_i^2}{m_i}.$$

从总的残差平方和 SS_E 中减去 SS_e 为模型 (9.2) 中的其他因素的影响误差, 记为 SS_{Me} , 即

$$SS_{Me} = SS_E - SS_e = \sum_{i=1}^k \frac{T_i^2}{m_i} - Y' \cdot X \cdot \hat{\beta},$$

称为模型的误差平方和, 其自由度分别为 $f_e = n - k, f_{Me} = k - m$.

在模型 (9.2) 为真的条件下, 可以得到

$$E(MS_e) = E\left(\frac{SS_e}{n-k}\right) = \sigma^2, E(MS_{Me}) = E\left(\frac{SS_{Me}}{k-m}\right) = \sigma^2,$$

且 SS_e 与 SS_{Me} 相互独立, 由 χ^2 分布的性质得

$$\frac{SS_e}{\sigma^2} \sim \chi^2(n-k), \frac{SS_{Me}}{\sigma^2} \sim \chi^2(k-m),$$

故

$$F = \frac{MS_{Me}}{MS_e} = \frac{SS_{Me}/k-m}{SS_e/n-k} \sim F(f_{Me}, f_e) = F(k-m, n-k),$$

即为拟合检验的统计量.

取一个显著水平 α (0.01 或 0.05), 对应地可查表得到 $F_{\alpha}(k-m, n-k)$, 用

数值计算 $F(k-m, n-k)$, 并与 $F_\alpha(k-m, n-k)$ 比较:

当 $F(k-m, n-k) < F_\alpha(k-m, n-k)$ 时, 则说明模型的拟合是好的, 是不显著的, 即模型的省略项造成的误差影响不大.

当 $F(k-m, n-k) > F_\alpha(k-m, n-k)$ 时, 则说明模型的拟合是不好的, 是显著的, 即模型的省略项造成的误差影响不可忽略, 需要增加新的变量.

现在的问题是: 如何增加新的变量? 就是下面的模型选择要解决的问题了.

9.3 回归模型的选择方法

由上面拟合性检验结果: $F(k-m, n-k) < (>) F_\alpha(k-m, n-k)$, 当“>”成立时拟合性检验是显著的, 即未考虑到的因素的影响不可忽略, 这就需要引入新的解释变量; 当“<”成立时拟合性检验是不显著的, 即模型中的解释变量已经足够了, 但并不是说模型中所有的变量都是必须的, 有可能有多余的变量, 应该从模型中去掉. 对于这两种情况, 如何去引入新的变量, 又如何去掉多余变量? 解决的方法就称为模型的选择.

模型选择的基本原则是: 既不遗漏一个重要的解释变量, 但也不把一个无用的解释变量保留在模型中. 如何考查一个解释变量在模型中的重要性呢? 主要是用它的偏回归平方和的大小来衡量, 何为偏回归平方和呢?

假设给定一组解释变量, 它的残差平方和为

$$SS_E = \hat{\varepsilon}^T \cdot \hat{\varepsilon} = (Y - X \cdot \hat{\beta})^T \cdot (Y - X \cdot \hat{\beta}) = Y^T \cdot Y - Y^T \cdot X \cdot \hat{\beta}.$$

9.3.1 去掉解释变量

假设把其中一个无用的解释变量去掉 (不妨去掉 $\varphi_m(u)$, 即 $\beta_m = 0$), 相应的可以计算残差平方和为 $SS_E^{(1)}$, 如果 $SS_E^{(1)} > SS_E$, 则称 $SS_E^{(1)} - SS_E$ 为解释变量 $\varphi_m(u)$ 的偏回归平方和, 它的大小是反映了 $\varphi_m(u)$ 在模型中贡献的大小, 即是衡量一个解释变量的重要性的定量指标. 究竟多大为重要 (需要保留), 多小为不重要 (可以去掉), 这就需要给出一个统计界限值.

不妨设要考查第 j 个解释变量 $\varphi_j(u)$ ($1 \leq j \leq m$) 的偏回归平方和.

如果已知 $\hat{\beta} = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$ 为回归系数的估计值, 相关矩阵为

$(X^T \cdot X)^{-1} = (C_{ij})$, 则可以证明: $\varphi_j(u)$ 的偏回归平方和为 $SS_E^{(j)} = \frac{\hat{\beta}_j^2}{C_{jj}}$ ($j = 1, 2,$

\dots, m), 其中 $\hat{\beta}_j$ 为 β_j 的估计值, C_{jj} 为相关矩阵的对角元素.

如果存在一个 k ($1 \leq k \leq m$) 使 $SS_E^{(k)} = \min_{1 \leq j \leq m} SS_E^{(j)}$, 即第 k 个解释变量 $\varphi_k(u)$ 在

模型中起的作用最小,能否把它去掉还要考查相应的 F 统计量

$$F = F(1, f_E) = \frac{SS_E^{(k)}}{MS_E},$$

其中 $MS_E = \frac{SS_E}{f_E}$ 为均值.

取一个显著水平 α , 对应地可查表得到 $F_\alpha(1, f_E)$, 用数值计算 $F(1, f_E)$, 并与 $F_\alpha(1, f_E)$ 比较:

当 $F(1, f_E) < F_\alpha(1, f_E)$ 时, 则称第 k 个解释变量 $\varphi_k(u)$ 是不显著的, 可以去掉.

当 $F(1, f_E) > F_\alpha(1, f_E)$ 时, 则称第 k 个解释变量 $\varphi_k(u)$ 是显著的, 不可以去掉. 并且其他的也都不能去掉.

注: 如果去掉一个变量后, 需要重新计算所有的偏回归平方和, 因为变量之间有相关性, 原来在 m 个变量中作用第二小的那个变量在 $m-1$ 个变量中其偏回归平方和不一定是最小的, 一般会发生变化.

9.3.2 增加解释变量

设要引进的变量为 $x_{m+1} = \varphi_{m+1}(u)$, 记为 x_{m+1} 在试验观测点 u_1, u_2, \dots, u_n 的值为

$$\bar{x}_{m+1} = (\varphi_{m+1}(u_1), \varphi_{m+1}(u_2), \dots, \varphi_{m+1}(u_n))^T,$$

则 m 个变量的回归系数的估计值取为

$$\hat{\beta}(\bar{x}_{m+1}) = (X^T \cdot X)^{-1} \cdot X^T \cdot \bar{x}_{m+1},$$

相应的残差平方和为

$$\begin{aligned} SS_E(\bar{x}_{m+1}, Y) &= (Y - X \cdot \hat{\beta}(\bar{x}_{m+1}))^T \cdot (\bar{x}_{m+1} - X \cdot \hat{\beta}(\bar{x}_{m+1})) \\ &= Y^T \cdot \bar{x}_{m+1} - Y^T \cdot X \cdot \hat{\beta}(\bar{x}_{m+1}). \end{aligned}$$

而

$$SS_E(\bar{x}_{m+1}, \bar{x}_{m+1}) = \bar{x}_{m+1}^T \cdot \bar{x}_{m+1} - \bar{x}_{m+1}^T \cdot X \cdot \hat{\beta}(\bar{x}_{m+1}),$$

则可以证明: x_{m+1} 的偏回归平方和为

$$SS_E^{(m+1)} = [SS_E(\bar{x}_{m+1}, Y)]^2 / SS_E(\bar{x}_{m+1}, \bar{x}_{m+1}).$$

$SS_E^{(m+1)}$ 的大小反映了 $\varphi_{m+1}(u)$ 对模型影响的大小, 即是衡量 $\varphi_{m+1}(u)$ 的作用的定量指标. 究竟 $SS_E^{(m+1)}$ 多大可以引进, 多小不需要引进呢? 这就需要建立统计量, 找出界限值.

假设 $m+1$ 个变量的残差平方和为 SS_E , 它比原 m 个变量的残差平方和 SS_E

要减少 $SS_E^{(m+1)}$, 即

$$SS_{\tilde{Y}} = SS_E - SS_E^{(m+1)},$$

相应的自由度为 $f_{\tilde{Y}} = f_E - f_E^{(m+1)} = n - m - 1$.

不妨设 $SS_E^{(m+1)}$ 是所有准备增加的变量中其偏回归平方和最大的一个, 它是否需要增加到模型中去, 要考查 F 统计量

$$F = F(1, f_{\tilde{Y}}) = \frac{MS_E^{(m+1)}}{MS_{\tilde{Y}}} = \frac{SS_E^{(m+1)}}{SS_{\tilde{Y}}/(n - m - 1)}.$$

取一个显著水平 α , 查表得 $F_\alpha(1, f_{\tilde{Y}})$, 计算得到 $F(1, f_{\tilde{Y}})$ 并比较二者大小. 如果 $F(1, f_{\tilde{Y}}) \geq F_\alpha(1, f_{\tilde{Y}})$, 则第 $m+1$ 个解释变量 $\varphi_{m+1}(u)$ 需要增加到模型中去, 否则无需增加, 而且也没有其他的变量需要增加了.

注: 在增加了 $\varphi_{m+1}(u)$ 以后, 可以继续上面的过程, 考查其他准备引入的变量中其偏回归平方和最大的那一个变量作为 $\varphi_{m+2}(u)$, 注意在 m 个变量中偏回归平方和第二大的那个变量在 $m+1$ 个中不一定是最大的, 这是因为变量有一定的相关性.

9.3.3 模型选择的一般方法

上面给出了在已知模型中剔除和增加解释变量的具体方法和步骤, 模型选择的一般方法如下:

(1) 淘汰法(向后法)

基本思想是: 把所有可选择的变量都放进模型中, 而后逐个做剔除检验, 直到不能剔除为止, 最后得到所选的模型.

(2) 纳新法(向前法)

基本思想是: 先少选取几个变量进入模型中, 而后对其他的变量逐个地做引入模型的检验, 直到不能引入为止, 得到最后的模型.

(3) 逐步回归法(吐故纳新法)

基本思想是: 结合上面的两种方法.

9.4 回归模型的正交化设计方法

由上面的讨论我们可以知道, 因为模型的解释变量之间有很复杂的相关性, 使回归系数的估计、模型的选择都带来很多的麻烦, 为了简化计算, 借助正交函数系可使问题简化.

9.4.1 正交的概念

设 $\varphi_1(u), \varphi_2(u), \dots, \varphi_m(u)$ 是 m 个解释变量, 如果对于 u_1, u_2, \dots, u_n 满足

$$(1) \sum_{i=1}^n \varphi_k^2(u_i) \neq 0 \quad (k = 1, 2, \dots, m);$$

$$(2) \sum_{i=1}^n \varphi_k(u_i) \varphi_j(u_i) = 0 \quad (j \neq k),$$

则称 $\varphi_1(u), \varphi_2(u), \dots, \varphi_m(u)$ 是正交的 ($m \leq n$).

如何构造正交函数系呢? 通常情况下, 正交函数都为正交多项式, 首先对于一维回归变量 u 来说明构造正交多项式的方法.

设有点列 u_1, u_2, \dots, u_n , 取 $\varphi_1(u) = 1, \varphi_2(u) = u - \bar{u}$, 其中 $\bar{u} = \frac{1}{n} \sum_{i=1}^n u_i$.

假设已做出了 k ($k \geq 1$) 阶正交多项式 $\varphi_1(u), \varphi_2(u), \dots, \varphi_{k+1}(u)$, 则第 $k+1$ 阶正交多项式为 $\varphi_{k+2}(u) = (u - a_{k+1})\varphi_{k+1}(u) - b_k\varphi_k(u)$, 其中

$$a_{k+1} = \frac{\sum_{i=1}^n u_i \varphi_{k+1}(u_i)}{\sum_{i=1}^n \varphi_{k+1}^2(u_i)}, \quad b_k = \frac{\sum_{i=1}^n \varphi_{k+1}^2(u_i)}{\sum_{i=1}^n \varphi_k^2(u_i)} \quad (i \leq k \leq n).$$

由此可以构造出任意阶的正交多项式.

一般说来, 在多维的回归变量的点列上构造正交多项式是很复杂的, 现在的问题是能否找到一种方法可将任意一组解释变量正交化的方法? 这就是下面的克拉姆-施密特 (Gram-Schmidt) 正交化方法.

设 $\varphi_1(u), \varphi_2(u), \dots, \varphi_m(u)$ 是由 u_1, u_2, \dots, u_n 确定的一组线性无关的解释变量, 构造 $\psi_1(u), \psi_2(u), \dots, \psi_m(u)$ 如下:

$$\psi_1(u) = \varphi_1(u),$$

$$\psi_2(u) = \varphi_2(u) - b_{21}\psi_1(u),$$

$$\psi_3(u) = \varphi_3(u) - b_{32}\psi_2(u) - b_{31}\psi_1(u),$$

$$\vdots$$

$$\psi_k(u) = \varphi_k(u) - b_{kk-1}\psi_{k-1}(u) - \dots - b_{k1}\psi_1(u),$$

$$\vdots$$

$$\psi_m(u) = \varphi_m(u) - b_{mm-1}\psi_{m-1}(u) - \dots - b_{m1}\psi_1(u),$$

$$\text{其中 } b_{kj} = \frac{\sum_{i=1}^n \varphi_k(u_i) \varphi_j(u_i)}{\sum_{i=1}^n \psi_j^2(u_i)}, \quad k = 2, 3, \dots, m; \quad j = 1, 2, \dots, k-1.$$

相当于对 $\varphi_1(u), \varphi_2(u), \dots, \varphi_m(u)$ 做了一个满秩变换, 可以验证 $\psi_1(u), \psi_2(u), \dots, \psi_m(u)$ 是在点列 u_1, u_2, \dots, u_n 上的正交的解釋变量。

9.4.2 正交性在建模中的应用

假设 $\varphi_1(u), \varphi_2(u), \dots, \varphi_m(u)$ 是 u_1, u_2, \dots, u_n 上的正交解釋变量, 建立模型如下:

$$\eta(u) = \beta_1 \varphi_1(u) + \beta_2 \varphi_2(u) + \dots + \beta_m \varphi_m(u).$$

又假设由 u_1, u_2, \dots, u_n 对应的观测值为 y_1, y_2, \dots, y_n , 则利用正交性可得回归系数的最小二乘估计值为

$$\hat{\beta}_k = \frac{\sum_{i=1}^n y_i \varphi_k(u_i)}{\sum_{i=1}^n \varphi_k^2(u_i)}, \quad k = 1, 2, \dots, m. \quad (9.4)$$

第 k 个解釋变量的偏回归平方和为

$$SS_E^{(k)} = \hat{\beta}_k^2 \left(\sum_{i=1}^n \varphi_k^2(u_i) \right) = \frac{\left(\sum_{i=1}^n y_i \varphi_k(u_i) \right)^2}{\sum_{i=1}^n \varphi_k^2(u_i)}, \quad (k = 1, 2, \dots, m). \quad (9.5)$$

残差平方和为

$$SS_E = \sum_{i=1}^n y_i^2 - \sum_{k=1}^m SS_E^{(k)}. \quad (9.6)$$

由此可以大大地简化计算, 而且在模型选择的检验中, 剔除变量或引入变量后其余变量的回归系数和偏回归平方和的值不改变 (因为它与变量个数无关)。

9.5 多重共线性与有偏估计方法

9.5.1 多重共线性

多重共线性是实际中由于解釋变量之间的相关性引起使正规方程组具有病态, 其解不稳定的一种现象。对于模型设计矩阵

$$X = \begin{bmatrix} \varphi_1(u_1) & \varphi_2(u_1) & \cdots & \varphi_m(u_1) \\ \varphi_1(u_2) & \varphi_2(u_2) & \cdots & \varphi_m(u_2) \\ \vdots & \vdots & & \vdots \\ \varphi_1(u_n) & \varphi_2(u_n) & \cdots & \varphi_m(u_n) \end{bmatrix},$$

一般假设 $r(X) = m$, 即为满秩矩阵, 则正规方程组的系数矩阵 $X^T \cdot X$ 为满秩阵, 亦即 $r(X^T \cdot X) = m$. 如果用 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ 表示 $X^T \cdot X$ 的 m 个特征根, 且当 $|X^T \cdot X| = \lambda_1 \lambda_2 \dots \lambda_m$ (行列式) 很小, 即至少有一个特征根接近于 0 (例如 λ_m 接近于 0), 但不等于 0 时, 则使正规方程组

$$X^T \cdot X \cdot \hat{\beta} = X^T \cdot Y$$

变成一种病态方程.

虽然 $\hat{\beta}$ 是 β 的无偏估计, 即 $E(\hat{\beta}) = \beta$, 但其均方误差

$$E(\hat{\beta} - \beta)^T (\hat{\beta} - \beta) = \sigma^2 \left(\frac{1}{\lambda_1} + \frac{1}{\lambda_2} + \dots + \frac{1}{\lambda_m} \right)$$

充分的大, 即使 β 的估计值 $\hat{\beta}$ 的误差太大, 无实用价值. 此时称 m 个解释变量之间具有多重共线性, 即也就是说设计矩阵 X 的列向量之间有近似的线性关系, 但非绝对的线性关系. 衡量多重共线性程度的量用

$$k = \frac{\text{最大特征根}}{\text{最小特征根}} = \frac{\lambda_1}{\lambda_m}$$

来表示,

- (1) 当 $k \leq 100$ 时, 则不存在多重共线性;
- (2) 当 $100 \leq k \leq 1\,000$ 时, 则存在较强的多重共线性;
- (3) 当 $k > 1\,000$ 时, 则存在严重的多重共线性.

思考题: 如何解决多重共线性的情况呢?

9.5.2 回归系数的有偏估计

在模型具有多重共线性时, 可使回归系数的估计值的偏差增大, 即均方差变大, 使估计不稳定, 为解决这个问题采用有偏估计法可以减少均方误差.

1. 岭估计法

基本思想: 由于 $|X^T \cdot X| \approx 0$, 使正规方程出现病态, 对一个常数 $k > 0$, 则有

$$|X^T \cdot X + kI| > |X^T \cdot X|.$$

用 $(X^T \cdot X + kI)\hat{\beta} = X^T \cdot Y$ 代替正规方程组, 其解为

$$\hat{\beta}(k) = (X^T \cdot X + kI)^{-1} \cdot X^T \cdot Y.$$

实际中, 根据需要可适当调整 k 的大小.

2. 主成份估计法

基本思想: 因为 $X^T \cdot X$ 是对称正定阵, 由代数的知识可知, 存在正交阵

$P(P^T P = I)$ 使

$$P^T \cdot X^T \cdot X \cdot P = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_m \end{bmatrix},$$

其中 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ 为 $X^T \cdot X$ 的 m 个特征根. 在有多重共线性时, 某些特征根近似为 0, 特征根的大小反映了对应的变量对模型影响的大小, 此方法就是忽略小的, 保留大的.

9.6 沼气的生成问题

9.6.1 问题的提出

沼气的主要成份为甲烷, 它是由含纤维素的有机物质在隔绝空气的情况下受到细菌分解作用所产生的一种有毒易燃气体. 在我国农村广泛地利用沼气池生成沼气, 作为一种卫生快捷的燃料, 一般是用植物结秆残体在保持一定湿度和温度的条件下, 并与空气隔绝一段时间后经自然分解而成. 试验证明, 如果适当地加入一些有机肥料作为发酵剂, 则可以加快沼气的形成. 下面是在一个确定沼气池中加入相同数量的同质植物结秆, 加入不同数量的水 (W) 和有机肥 (F) 后形成沼气的时间 (T) 对比数据 (如表 9-1), 请根据这些试验数据分析研究沼气形成的时间与水和有机肥料之间的关系, 并由此关系讨论最佳的配料方案.

表 9-1 W, F 和 T 的试验数据

| 试验次数 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| $W(\text{kg})$ | 300 | 400 | 500 | 300 | 400 | 500 | 300 | 400 | 500 |
| $F(\text{kg})$ | 200 | 200 | 200 | 250 | 250 | 250 | 300 | 300 | 300 |
| $T(\text{h})$ | 77 | 68 | 59 | 66 | 62 | 52 | 59 | 55 | 50 |

9.6.2 模型的假设与分析

• 模型假设

(1) 设试验数据是在相同的试验条件下进行的, 即沼气池大小形状相同, 结秆和有机肥料相同, 其自身的含水量也相同;

(2) 在此不考虑环境温度的影响, 虽然在同等条件下高温可以促使沼气的形成, 但实际中的环境温度一般是不可控的, 于是我们认为总是在一定的适宜温

度范围内,温度因素对形成沼气的时影响不大;

(3) 每次实验是独立进行的,且 W 、 F 和 T 的试验值是准确的.

• 模型的分析

根据实际中沼气的自然形成的原理和有关的常识,我们知道在同等条件下,水分和肥料各自都对沼气的形成有一定的促进作用,而且二者之间也有一定的交互效应,即二者用量的多少不同其效果是不同的. 即沼气的形成时间不仅与水和肥料的用量有关,而且还与二者的交互作用有关. 因此,一般认为沼气形成时间 T 的长短应该是加水量 W 和肥料用量 F 的二次多项式函数,为此,我们可以采用线性回归方法来研究它们之间的关系式.

9.6.3 模型的建立与求解

为了便于对问题的描述,我们不直接将沼气形成的时间 T 表示成 W, F 的函数,根据试验数据的分布情况,在这里引入两个新的变量

$$u_1 = \frac{W - 400}{100}, u_2 = \frac{F - 250}{50}.$$

为此可以将时间 T 表示成 u_1, u_2 的二次多项式函数.

我们首先来构造正交的多项式,由试验数据(如表 9-1)可得相应的新数据如表 9-2.

表 9-2 W, F 的试验数据转换为 u_1, u_2 的数据

| 试验次数 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | \bar{u}_i |
|-------|----|----|----|----|---|---|----|---|---|-------------|
| u_1 | -1 | 0 | 1 | -1 | 0 | 1 | -1 | 0 | 1 | 0 |
| u_2 | -1 | -1 | -1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |

实际上可以证明:

$$\varphi_1(u) = 1, \varphi_2(u) = u_1, \varphi_3(u) = u_1^2 - \frac{2}{3}, \varphi_4(u) = u_2,$$

$$\varphi_5(u) = u_2^2 - \frac{2}{3}, \varphi_6(u) = u_1 u_2$$

在 9 个试验点上是正交的,其中向量 $u = (u_1, u_2)^T$.

于是有回归模型的一般形式为

$$\begin{aligned} T(u) &= \beta_1 \varphi_1(u) + \beta_2 \varphi_2(u) + \beta_3 \varphi_3(u) + \beta_4 \varphi_4(u) + \beta_5 \varphi_5(u) + \beta_6 \varphi_6(u) \\ &= \beta_1 + \beta_2 u_1 + \beta_3 \left(u_1^2 - \frac{2}{3} \right) + \beta_4 u_2 + \beta_5 \left(u_2^2 - \frac{2}{3} \right) + \beta_6 u_1 u_2. \end{aligned}$$

在这里可以用最小二乘法求出所有的回归系数 $\beta_i (i=1, 2, \dots, 6)$. 实际上, 根据其正交性, 将表 9-2 中数据代入 (9.4) 式中计算可得回归系数的估计值为

$$\hat{\beta}_1 = \frac{549}{9} = 61, \hat{\beta}_2 = -\frac{41}{6} \approx -6.83, \hat{\beta}_3 = -\frac{7}{6} \approx -1.17,$$

$$\hat{\beta}_4 = -\frac{20}{3} \approx -6.67, \hat{\beta}_5 = \frac{4}{3} \approx 1.33, \hat{\beta}_6 = \frac{9}{4} = 2.25.$$

由公式 (9.5), 各个变量的偏回归平方和为

$$SS_E^{(1)} = 33\,367, SS_E^{(2)} = 280.17, SS_E^{(3)} = 2.72,$$

$$SS_E^{(4)} = 266.67, SS_E^{(5)} = 3.56, SS_E^{(6)} = 20.25.$$

又由公式 (9.6), 总残差平方和为

$$SS_E = \sum_{i=1}^9 y_i^2 - \sum_{k=1}^6 SS_E^{(k)} \approx 33\,944 - 33\,940.47 = 3.53,$$

而且其自由度为 $f_E = 9 - 6 = 3$. 在所有偏回归平方和中最小的是 $SS_E^{(3)} = 2.72$, 对应的解释变量为 $\varphi_3(u) = u_1^2 - \frac{2}{3}$, 它是否要从模型中去掉, 需要做进一步的显著性检验.

由于 $MS_E^{(3)} = SS_E^{(3)} = 2.72, MS_E = \frac{SS_E}{f_E} = 1.177$, 则 F 统计量为

$$F(1, 3) = \frac{MS_E^{(3)}}{MS_E} = \frac{2.72}{1.177} = 2.31.$$

取显著水平 $\alpha = 0.05$ 时, 查表得 $F_\alpha(1, 3) = 10.1$, 即 $F(1, 3) < F_\alpha(1, 3)$, 于是 $\varphi_3(u) = u_1^2 - \frac{2}{3}$ 在模型中的作用是不显著的, 可以将此项从模型中剔除. 而后应将相应的偏回归平方和加入到总残差平方和中 (这是因为模型是由正交变量构成的, 所以可以直接求和), 即为 $SS_E = 3.53 + 2.72 = 6.25$, 自由度为 $f_E = 4$, 均值为 $MS_E = 1.5625$.

下面要进一步考查偏回归平方和次小的解释变量的显著性. 显然, 次小的是 $SS_E^{(5)} = 3.56$, 对应的解释变量为 $\varphi_5(u) = u_2^2 - \frac{2}{3}$, 类似的可以计算

$$F(1, 4) = \frac{MS_E^{(5)}}{MS_E} = \frac{3.56}{1.5625} = 2.2784.$$

而对于显著水平 $\alpha = 0.05$, 查表得 $F_\alpha(1, 4) = 7.71$, 即 $F(1, 4) < F_\alpha(1, 4)$, 于是 $\varphi_5(u)$ 在模型中的作用还是不显著的. 可以将此项从模型中剔除. 而后应将相应

的偏回归平方和加入到总残差平方和中去,即为 $SS_E = 6.25 + 3.56 = 9.81$, 自由度为 $f_E = 5$, 均值为 $MS_E = 1.962$.

再来考查偏回归平方和第三小的解释变量的显著性. 显然是 $SS_E^{(6)} = 20.25$, 对应的解释变量为 $\varphi_6(u) = u_1 u_2$, 类似的可以计算

$$F(1, 5) = \frac{MS_E^{(6)}}{MS_E} = \frac{20.25}{1.962} = 10.32.$$

对于显著水平 $\alpha = 0.05$, 查表得 $F_\alpha(1, 5) = 6.61$, 即 $F(1, 5) > F_\alpha(1, 5)$, 于是 $\varphi_6(u)$ 在模型中的作用还是显著的, 即正好反映出了水 (W) 和肥料 (F) 对生成沼气的交互作用. 到此为止, 模型中也没有可以剔除的变量了, 故此, 我们最后确定的回归模型为

$$T(u) = 61 - \frac{41}{6}u_1 - \frac{20}{3}u_2 + \frac{9}{4}u_1 u_2. \quad (9.7)$$

将 $u_1 = \frac{W-400}{100}$, $u_2 = \frac{F-250}{50}$ 代入上可以得沼气的生成时间 T 与水 W 和肥料 F 的函数关系式.

由 (9.7) 式可得: 当 $u_1 = \frac{80}{27} \approx 2.963$, $u_2 = 3$, 即 $W = 696$ kg 和 $F = 400$ kg 时, 生成沼气的时间有最小值 $T = 40$ h.

9.7 参考案例与参考文献

1. 参考案例

- (1) 土豆的生长施肥问题——文献[1]:38—40
- (2) 教学评估问题——文献[2]:322—326
- (3) 气象观测站问题——文献[3]:320—329
- (4) 投资额与生产总值和物价指数问题——文献[2]:316—322
- (5) 施肥效果分析问题——文献[5]:1

2. 参考文献

- [1] 叶其孝. 大学生数学建模竞赛辅导教材(一). 长沙: 湖南教育出版社, 1997
- [2] 姜启源. 数学模型. 第三版. 北京: 高等教育出版社, 2003
- [3] 寿纪麟. 数学建模方法与范例. 西安: 西安交通大学出版社, 1993
- [4] 《数学模型》编写组. 数学模型. 广州: 华南理工大学出版社, 2003
- [5] 全国大学生数学建模竞赛组委会. 全国大学生数学建模竞赛优秀论文汇编. 北京: 中国物价出版社, 2002

- [6] 赵静,但琦等. 数学建模与数学实验. 北京:高等教育出版社,2000
- [7] 南京地区工科院校数学建模与工业数学讨论班. 数学建模与实验. 南京:河海大学出版社,1996
- [8] 中山大学. 概率论及数理统计. 北京:高等教育出版社,1984
- [9] 陈魁. 应用概率统计. 北京:清华大学出版社,2000

第十章 线性规划方法

实际中所研究的优化问题,一般都是要求使问题的某一项指标“最优”的方案,这里的“最优”包括“最好”、“最大”、“最小”、“最高”、“最低”、“最多”、“最少”等等,这类问题统称为**最优化问题**,解决这类问题的最常用方法就是线性规划方法.目前,因为线性规划有着非常完备的理论基础和有效的求解方法,所以线性规划在实际中应用是十分广泛的,譬如像合理地分配、使用有限的资源(经济、人力、物资等),使能够获得“最优效益”等问题.

10.1 线性规划的模型

10.1.1 问题的引入

设某企业现有 m 种资源 $A_i (i = 1, 2, \dots, m)$ 用于生产 n 种产品 $B_j (j = 1, 2, \dots, n)$, 每种资源的拥有量和每种产品所消耗的资源量以及单位产品的利润如表 10-1, 试问如何安排生产计划使得该企业获利最大?

表 10-1

| 产品 资源 | B_1 | B_2 | ... | B_n | 总量 |
|----------|----------|----------|-----|----------|----------|
| A_1 | a_{11} | a_{12} | ... | a_{1n} | b_1 |
| A_2 | a_{21} | a_{22} | ... | a_{2n} | b_2 |
| \vdots | \vdots | \vdots | | \vdots | \vdots |
| A_m | a_{m1} | a_{m2} | ... | a_{mn} | b_m |
| 利润 | c_1 | c_2 | ... | c_n | |

建立数学模型:

设产品 B_j 产量为 $x_j (j = 1, 2, \dots, n)$, 称之为**决策变量**, 所得的利润为 z , 则要解决的问题的目标是使得(总利润)函数 $z = \sum_{j=1}^n c_j x_j$ 有最大值. 决策变量所受的

约束条件为

$$\begin{cases} \sum_{j=1}^n a_{ij}x_j \leq b_i & (i = 1, 2, \dots, m), \\ x_j \geq 0 & (j = 1, 2, \dots, n). \end{cases}$$

于是问题可归结为求目标函数在约束条件下的最大值问题. 显然的目标函数和约束条件都是决策变量的线性函数, 即有下面的线性规划模型

$$\text{目标函数: } \max z = \sum_{j=1}^n c_j x_j,$$

$$\text{约束条件: } \begin{cases} \sum_{j=1}^n a_{ij}x_j \leq b_i & (i = 1, 2, \dots, m), \\ x_j \geq 0 & (j = 1, 2, \dots, n). \end{cases} \quad (10.1)$$

一般地, 如果问题的目标函数和约束条件关于决策变量都是线性的, 则称该问题为线性规划问题, 其模型称为线性规划模型.

10.1.2 线性规划模型的一般形式

线性规划模型的一般形式为

$$\begin{aligned} \max(\min) z &= \sum_{j=1}^n c_j x_j, \\ \text{s. t. } \begin{cases} \sum_{j=1}^n a_{ij}x_j \leq (\geq, =) b_i & (i = 1, 2, \dots, m), \\ x_j \geq 0 & (j = 1, 2, \dots, n). \end{cases} \end{aligned}$$

也可表示为矩阵形式

$$\begin{aligned} \max(\min) z &= C \cdot X, \\ \text{s. t. } \begin{cases} A \cdot X \leq (\geq, =) b, \\ X \geq 0. \end{cases} \end{aligned}$$

向量形式

$$\begin{aligned} \max(\min) z &= C \cdot X, \\ \text{s. t. } \begin{cases} \sum_{j=1}^n P_j x_j \leq (\geq, =) b, \\ X \geq 0, \end{cases} \end{aligned}$$

其中 $C = (c_1, c_2, \dots, c_n)$ 称其为目标函数的系数向量; $X = (x_1, x_2, \dots, x_n)^T$ 称其为决策向量; $b = (b_1, b_2, \dots, b_m)^T$ 称其为约束方程组的常数向量; $A = (a_{ij})_{m \times n}$ 称其为约束方程组的系数矩阵; $P_j = (a_{1j}, a_{2j}, \dots, a_{mj})^T (j = 1, 2, \dots, n)$ 称其为约束方程组的系数向量.

10.1.3 线性规划模型的标准形

线性规划模型的标准形规定为

$$\max z = C \cdot X \quad (10.2)$$

$$\text{s. t. } \begin{cases} A \cdot X = b \\ X \geq 0 \end{cases} \quad (10.3)$$

对于非标准形的线性规划模型都可以化为标准形,其方法如下:

- (1) 目标函数为最小化问题:令 $z' = -z$, 则 $\max z' = -\min z = -C \cdot X$;
- (2) 约束条件为不等式:对于不等号“ \leq ”的约束条件,则可在“ \leq ”的左端加上(或减去)一个非负变量(称为松弛变量)使其变为等式;
- (3) 对于无约束的决策变量:譬如 $x \in (-\infty, +\infty)$, 则令 $x = x' - x''$, 使得 $x', x'' \geq 0$, 代入模型即可.

10.2 线性规划解的概念与理论

10.2.1 线性规划解的概念

(1) 解:称满足约束条件((10.3)式)的解 $X = (x_1, x_2, \dots, x_n)^T$ 为线性规划问题的可行解;可行解的全体构成的集合称为可行域,记为 D ;使目标函数((10.2)式)达到最大的可行解称为最优解.

(2) 基:设系数矩阵 $A = (a_{ij})_{m \times n}$ 的秩为 m , 则称 A 的某个 $m \times m$ 阶非奇异子矩阵 B ($\det B \neq 0$) 为线性规划问题的一个基.不妨设 $B = (a_{ij})_{m \times m} = (P_1, P_2, \dots, P_m)$, 则称向量 $P_j = (a_{1j}, a_{2j}, \dots, a_{mj})^T$ ($j=1, 2, \dots, m$) 为基向量,其他的称为非基向量;与基向量对应的决策变量 x_j ($j=1, 2, \dots, m$) 称为基变量,其他的变量称为非基变量.

(3) 基解:设问题的基为 $B = (a_{ij})_{m \times m} = (P_1, P_2, \dots, P_m)$, 将约束方程组变为

$$\sum_{j=1}^m P_j x_j = b - \sum_{j=m+1}^n P_j x_j. \quad (10.4)$$

在方程组(10.4)的解中令 $x_j = 0$ ($j=m+1, \dots, n$), 则称解向量 $X = (x_1, x_2, \dots, x_m, 0, \dots, 0)^T$ 为线性规划问题的基解.

(4) 基可行解:满足非负约束条件的基解称为基可行解.

(5) 可行基:对应于基可行解的基称为可行基.

10.2.2 线性规划解的基本理论

定理 10.1 如果线性规划问题(10.2), (10.3)存在可行域, 则其可行域

$$D = \left\{ X \mid \sum_{j=1}^n P_j x_j = b, x_j \geq 0 \right\}$$

是凸集.

定理 10.2 线性规划问题(10.2),(10.3)的任一个基可行解 X 必对应于可行域 D 的一个顶点.

定理 10.3 (1) 如果线性规划问题(10.2),(10.3)的可行域有界,则问题的最优解一定在可行域的顶点上达到;

(2) 如果线性规划问题(10.2),(10.3)的可行域有无界解,则问题可能无最优解;若有最优解也一定在可行域的某个顶点上达到.

10.3 线性规划的求解方法

根据线性规划的解的概念和基本理论,求解线性规划可采用下面的方法:求一个基可行解;检查该基可行解是否为最优解;如果不是,则设法再求另一个没有检查过的基可行解,如此进行下去,直到得到某一个基可行解为最优解为止.现在要解决的问题是:如何求出第一个基可行解?如何判断基可行解是否为最优解?如何由一个基可行解过渡到另一个基可行解?解决这些问题的方法称为单纯形法.其基本步骤如下:

10.3.1 初始基可行解的确定

如果线性规划问题为标准形(即约束方程全为等式),则从系数矩阵 $A = (a_{ij})_{m \times n}$ 中观察总可以得到一个 m 阶单位阵 I_m . 如果问题的约束条件的不等号均为“ \leq ”,则引入 m 个松弛变量,可化为标准形,并将变量重新排序编号,即可得到一个 m 阶单位阵 I_m ;如果问题的约束条件的不等号为“ \geq ”和“ $=$ ”,则首先引入松弛变量化为标准形,再通过人工变量法总能得到一个 m 阶单位阵 I_m . 综上所述,取如上 m 阶单位阵 I_m 为初始可行基,即 $B = I_m$,将相应的约束方程组变为

$$x_i = b_i - a_{im+1}x_{m+1} - \cdots - a_{in}x_n, i = 1, 2, \cdots, m.$$

令 $x_j = 0$ ($j = m+1, \cdots, n$),则可得一个初始基可行解

$$X^{(0)} = (x_1^{(0)}, x_2^{(0)}, \cdots, x_m^{(0)}, 0, \cdots, 0)^T = (b_1, b_2, \cdots, b_m, 0, \cdots, 0)^T.$$

10.3.2 寻找另一个基可行解

当一个基可行解不是最优解或不能判断时,需要过渡到另一个基可行解,即从基可行解 $X^{(0)} = (x_1^{(0)}, x_2^{(0)}, \cdots, x_m^{(0)}, 0, \cdots, 0)^T$ 对应的可行基 $B = (P_1, P_2, \cdots, P_m)$ 中替换一个列向量,并与原向量组线性无关. 譬如用非基变量 P_{m+t} ($1 \leq t \leq$

$n-m$) 替换基变量 $P_l (1 \leq l \leq m)$, 就可得到一个新的可行基 $B_l = (P_1, \dots, P_{l-1}, P_{m+t}, P_{l+1}, \dots, P_m)$, 从而可以求出一个新的基可行解 $X^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_m^{(1)}, 0, \dots, 0)^T$, 其方法称为基变换法. 事实上

$$x_i^{(1)} = \begin{cases} x_i^{(0)} - \theta \beta_{i,m+t}, & i \neq l, \quad (i = 1, 2, \dots, m, \\ \theta, & i = l \quad (1 \leq l \leq m, 1 \leq t \leq n-m), \end{cases}$$

$$\text{其中 } \theta = \frac{x_l^{(0)}}{\beta_{l,m+t}} = \min_{1 \leq i \leq m} \left\{ \frac{x_i^{(0)}}{\beta_{i,m+t}} \mid \beta_{i,m+t} > 0 \right\}, \quad P_{m+t} = \sum_{i=1}^m \beta_{i,m+t} P_i.$$

如果 $X^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_m^{(1)}, 0, \dots, 0)^T$ 仍不是最优解, 则可以重复利用这种方法, 直到最优解为止.

10.3.3 最优性检验的方法

假设要检验基可行解 $X^{(1)} = (x_1^{(1)}, x_2^{(1)}, \dots, x_m^{(1)}, 0, \dots, 0)^T = (b'_1, b'_2, \dots, b'_m, 0, \dots, 0)^T$ 的最优性. 由约束方程组对任意的 $X = (x_1, x_2, \dots, x_n)^T$ 有

$$x_i = b'_i - \sum_{j=m+1}^n a'_{ij} x_j, \quad i = 1, 2, \dots, m.$$

将基可行解 $X^{(1)}$ 和任意的 $X = (x_1, x_2, \dots, x_n)^T$ 分别代入目标函数得

$$\begin{aligned} z^{(0)} &= \sum_{i=1}^m c_i x_i^{(1)} = \sum_{i=1}^m c_i b'_i, \\ z^{(1)} &= \sum_{i=1}^m c_i x_i = \sum_{i=1}^m c_i x_i + \sum_{i=m+1}^n c_i x_i \\ &= \sum_{i=1}^m c_i \left(b'_i - \sum_{j=m+1}^n a'_{ij} x_j \right) + \sum_{j=m+1}^n c_j x_j \\ &= \sum_{i=1}^m c_i b'_i + \sum_{j=m+1}^n \left(c_j - \sum_{i=1}^m c_i a'_{ij} \right) x_j \\ &= z^{(0)} + \sum_{j=m+1}^n (c_j - z_j) x_j, \end{aligned}$$

其中 $z_j = \sum_{i=1}^m c_i a'_{ij} \quad (j = m+1, \dots, n)$. 记 $\sigma_j = c_j - z_j \quad (j = m+1, \dots, n)$, 则

$$z^{(1)} = z^{(0)} + \sum_{j=m+1}^n \sigma_j x_j. \quad (10.5)$$

注: 当 $\sigma_j > 0$ 时就有 $z^{(1)} \geq z^{(0)}$; 当 $\sigma_j \leq 0$ 时就有 $z^{(1)} \leq z^{(0)}$. 为此, $\sigma_j = c_j - z_j$ 的符号是判别 $X^{(1)}$ 是否为最优解的关键所在, 故称之为检验数. 于是由 (10.5) 式可以有下面的结论:

- (1) 如果 $\sigma_j \leq 0 \quad (j = m+1, \dots, n)$, 则 $X^{(1)}$ 是问题的最优解, 最优值为 $z^{(0)}$;
- (2) 如果 $\sigma_j \leq 0 \quad (j = m+1, \dots, n)$ 且至少存在一个 $\sigma_{m+k} = 0 \quad (1 \leq k \leq n-m)$

m), 则问题有无穷多个最优解, $X^{(1)}$ 是其中之一, 最优值为 $z^{(0)}$;

(3) 如果 $\sigma_j < 0$ ($j = m+1, \dots, n$), 则 $X^{(1)}$ 是问题的惟一的最优解, 最优值为 $z^{(0)}$;

(4) 如果存在某个检验数 $\sigma_{m+k} > 0$ ($1 \leq k \leq n-m$), 并且对应的系数向量 P_{m+k} 的各分量 $a_{i,m+k} \leq 0$ ($i = 1, 2, \dots, m$), 则问题具有无界解(即无最优解).

10.4 线性规划的对偶问题

10.4.1 对偶问题的提出

将 10.1.1 节中提出的实际问题从相反的角度提出: 假设有 B 企业要将 A 企业的资源和生产权全部收买过来, 问题是 B 企业至少应付多少代价, 才能使 A 企业愿意转让所有资源和生产权?

事实上, 要让 A 转让的条件是: 对同等数量的资源出让的代价不应低于 A 企业自己生产的产值, 即若用 y_i 表示 B 收买 A 的一个单位第 i 种资源时付出的代价, 则 A 出让生产一个单位第 j 种产品资源的价值不应低于生产一个单位第 j 种产品的产值 c_j 元, 即

$$\sum_{i=1}^m a_{ij} y_i \geq c_j \quad (j = 1, 2, \dots, n).$$

对 B 企业, 希望花最小的代价将 A 的所有资源及生产权收买过来, 即问题为

$$\begin{aligned} \min w &= \sum_{i=1}^m b_i y_i, \\ \begin{cases} \sum_{i=1}^m a_{ij} y_i \geq c_j & (j = 1, 2, \dots, n), \\ y_i \geq 0 & (i = 1, 2, \dots, m) \end{cases} \end{aligned} \quad (10.6)$$

或

$$\{\min w = Y \cdot b \mid Y \cdot A \geq C, Y \geq 0\}.$$

这也是一个线性规划问题, 问题 (10.6) 称为问题 (10.1) 的**对偶问题**, 问题 (10.1) 称为对偶问题 (10.6) 的**原问题**, 即两者为相互对偶的问题.

10.4.2 原问题与对偶问题的关系

原问题与对偶问题的关系如表 10-2. 正面看是原问题, 顺时针旋转 90 度看是对偶问题. 如果约束条件中的不等号反向或为等式, 对偶问题的变化情况如表 10-3.

表 10-2 原问题与对偶问题的关系

| x_j | x_1 | x_2 | \cdots | x_n | 原关系 | $\min w$ |
|----------|----------|----------|----------|----------|-------------------|----------|
| y_i | | | | | | |
| y_1 | a_{11} | a_{12} | \cdots | a_{1n} | \leq | b_1 |
| y_2 | a_{21} | a_{22} | \cdots | a_{2n} | \leq | b_2 |
| \vdots | \vdots | \vdots | | \vdots | \vdots | \vdots |
| y_m | a_{m1} | a_{m2} | \cdots | a_{mn} | \leq | b_m |
| 对偶关系 | \geq | \geq | \cdots | \geq | $\max z = \min w$ | |
| $\max z$ | c_1 | c_2 | \cdots | c_n | | |

表 10-3 对偶问题的变化情况

| 原问题 ($\max z$) | | 对偶问题 ($\min w$) | |
|------------------|----------|-------------------|----------|
| 约束条件 (m 个) | \leq | 变量符号 (m 个) | ≥ 0 |
| | \geq | | ≤ 0 |
| | $=$ | | 无约束 |
| 变量符号 (n 个) | ≥ 0 | 约束条件 (n 个) | \geq |
| | ≤ 0 | | \leq |
| | 无约束 | | $=$ |

注：使用此表时总是视为最大化问题为原问题，最小化问题视为对偶问题，否则会容易出错。

设线性规划的原问题为 $\{\max z = C \cdot X \mid AX \leq b, X \geq 0\}$ ，相应的对偶问题为 $\{\min w = Y \cdot b \mid Y \cdot A \geq C, Y \geq 0\}$ ，则有如下性质：

- (1) 对偶问题的对偶问题是原问题；
- (2) 如果原问题(对偶问题)为无界解，则其对偶问题(原问题)无可行解，反之不然；
- (3) 设 \hat{X} 是原问题的可行解， \hat{Y} 是对偶问题的可行解，且 $C \cdot \hat{X} = \hat{Y} \cdot b$ ，则 \hat{X} 和 \hat{Y} 分别是原问题和对偶问题的最优解；
- (4) 如果原问题有最优解，则其对偶问题也一定有最优解，且有 $\max z = \min w$ 。

10.4.3 对偶单纯形法

根据对偶问题的性质和原问题与对偶问题的解之间的关系，原问题的检验数是对偶问题的基解，求解中通过若干步的迭代后，当原问题检验数为对偶问题的基可行解时，则也就得到了原问题和对偶问题的最优解。迭代中主要是根据检验数的符号判断是否得到了最优解。

对偶单纯形法的步骤：

- (1) 根据所给问题化为标准形，并写出相应的对偶问题。

注：无需引入人工变量，初始解可以不是可行解，在迭代的过程中可逐步靠近可行解，最后达到可行解，即为最优解。

- (2) 检验是否得到最优解：即检验 b 列数据 $(B^{-1}b)_i (i = 1, 2, \cdots, m)$ 和检验数 $\sigma_j (j = m+1, \cdots, n)$ 的符号：

如果 $(B^{-1}b)_i \geq 0$ 且 $\sigma_j \leq 0$ ，则已得到最优解，停止计算；

如果存在 $(B^{-1}b)_i < 0$ 且 $\sigma_j \leq 0$, 则进行下一步.

(3) 确定换出变量: 求 $\min_i \{ (B^{-1}b)_i \mid (B^{-1}b)_i < 0 \} = (B^{-1}b)_l$, 对应的基变量 $x_l (1 \leq l \leq m)$ 为换出变量.

(4) 确定换入变量: 检查 x_l 所在的行的各系数 $a_{lj} (j=1, 2, \dots, n)$ 的符号:

如果 $a_{lj} \geq 0 (j=1, 2, \dots, n)$, 则问题无可行解, 停止计算;

如果至少存在一个 $a_{lj} < 0$, 则计算 $\theta = \min_j \left\{ \frac{\sigma_j}{a_{lj}} \mid a_{lj} < 0 \right\} = \frac{\sigma_k}{a_{lk}} (m+1 \leq k \leq n)$,

对应的非基变量 x_k 为换入变量.

(5) 以 a_{lk} 为主元, 用初等变换法, 将系数矩阵中的 k 列元与 l 列元对换 (即将第 k 列中除 l 行元为 1 外, 其他都为 0), 即得到新的矩阵.

重复上面的 (2) — (5), 直到得到最优解为止.

对偶单纯形的特点:

(1) 原问题的初始解不需要是可行解, 因此, 不必引进人工变量, 使计算简化;

(2) 当变量的个数多于约束条件的个数时, 用对偶单纯形法可大大减少工作量, 因此, 当问题的变量个数少, 而约束条件个数多时, 可以将问题转化为对偶问题, 然后用对偶单纯形法求解;

(3) 在对偶单纯形法中, 找到一个可行的初始解较困难, 因此, 一般对偶单纯形法不单独使用, 多用于整数规划和灵敏度分析中.

10.5 线性规划的灵敏度分析

在线性规划模型 $\{\max z = C \cdot X \mid AX = b, X \geq 0\}$ 中, 我们总是假设 A, b, C 都是常数向量, 但实际中这些数值许多都是由试验或测量得到的试验值和预测值, 特别是在迭代计算中也都是近似值. 一般 A 表示工艺条件, b 表示资源条件, C 表示市场条件, 实际中多种原因都可能引起它们的变化. 现在的问题是: 这些系数在什么范围内变化时, 使线性规划问题最优解不变? 这就是灵敏度分析要研究的问题.

10.5.1 市场条件(价值系数) C 的变化分析

设 C 中的第 k 个元 c_k 发生变化, 即 $c'_k = c_k + \Delta c_k$, 其他不变, 问题是: 当 c_k 在什么范围变化时使问题的最优解不变?

(1) 若 c_k 是非基变量 x_k 的系数, 则对应的检验数为 $\sigma_k = c_k - C_B B^{-1} P_k$, 于是 $\sigma'_k = c'_k + \Delta c_k - C_B B^{-1} P_k$. 当 $\sigma'_k \leq 0$, 即 $\sigma \Delta c_k \leq C_B B^{-1} P_k - c_k = Y \cdot P_k - c_k$.

(2) 若 c_k 是基变量 x_k 的系数, 当 c_k 有改变量 Δc_k 时, 即 $c'_k = c_k + \Delta c_k$, 则 $C'_B = C_B + \Delta C_B$, $\Delta C_B = (0, 0, \dots, \Delta c_k, \dots, 0)$, 其相应的检验数为

$$\begin{aligned}\sigma' &= C - C'_B \cdot B^{-1} \cdot A = C - C_B \cdot B^{-1} \cdot A - \Delta C_B \cdot B^{-1} \cdot A \\ &= C - C_B \cdot B^{-1} \cdot A - (0, 0, \dots, \Delta c_k, \dots, 0) \cdot B^{-1} \cdot A,\end{aligned}$$

即

$$\sigma'_j = c_j - C_B B^{-1} P_j - \Delta c_k \bar{a}_{kj} \quad (j = 1, 2, \dots, n).$$

当 $\sigma'_j \leq 0$ 时, 问题的最优解不变, 故有 $c_j - C_B B^{-1} P_j - \Delta c_k \bar{a}_{kj} \leq 0$, 即 $\sigma_j - \Delta c_k \bar{a}_{kj} \leq 0$.

于是, 当 $\bar{a}_{kj} < 0$ 时, $\Delta c_k \leq \frac{\sigma_j}{\bar{a}_{kj}} \quad (j = 1, 2, \dots, n)$; 当 $\bar{a}_{kj} > 0$ 时, $\Delta c_k \geq \frac{\sigma_j}{\bar{a}_{kj}} \quad (j = 1, 2, \dots, n)$. 故 Δc_k 的允许变化范围为

$$\max_j \left\{ \frac{\sigma_j}{\bar{a}_{kj}} \mid \bar{a}_{kj} > 0 \right\} \leq \Delta c_k \leq \min_j \left\{ \frac{\sigma_j}{\bar{a}_{kj}} \mid \bar{a}_{kj} < 0 \right\}.$$

10.5.2 资源条件 b 变化的分析

设 b 中的第 k 个元 b_k 发生变化, 即 $b'_k = b_k + \Delta b_k$, 其他系数均不变, 则问题的解变化为 $X'_B = B^{-1}(b + \Delta b)$, $\Delta b = (0, 0, \dots, \Delta b_k, \dots, 0)^T$.

当 $X'_B \geq 0$ 时, 检验数不变, 则最优基(最优解对应的基)不变, 但最优解的值要发生变化. 下面考察 Δb_k 在什么范围变化时, 最优解变化不大.

因为新的最优解为 $X'_B = B^{-1}(b + \Delta b) = B^{-1}b + B^{-1}\Delta b$, 所以

$$\begin{aligned}B^{-1}\Delta b &= B^{-1} \cdot (0, 0, \dots, \Delta b_k, \dots, 0)^T \\ &= (\bar{a}_{1k}\Delta b_k, \dots, \bar{a}_{ik}\Delta b_k, \dots, \bar{a}_{mk}\Delta b_k)^T \\ &= \Delta b_k (\bar{a}_{1k}, \dots, \bar{a}_{ik}, \dots, \bar{a}_{mk})^T \quad (i = 1, 2, \dots, m),\end{aligned}$$

则最后求得的 b 列元为 $\bar{b}_i + \bar{a}_{ik}\Delta b_k \geq 0 \quad (i = 1, 2, \dots, m)$, 即 $\bar{a}_{ik}\Delta b_k \geq -\bar{b}_i \quad (i = 1, 2, \dots, m)$, 其中 \bar{b}_i 为 $B^{-1}b$ 的元.

注: 当 $\bar{a}_{ik} > 0$ 时, $\Delta b_k \geq -\frac{\bar{b}_i}{\bar{a}_{ik}} \quad (i = 1, 2, \dots, m)$;

当 $\bar{a}_{ik} < 0$ 时, $\Delta b_k \leq -\frac{\bar{b}_i}{\bar{a}_{ik}} \quad (i = 1, 2, \dots, m)$. 故 Δb_k 的允许变化范围为

$$\max_i \left\{ -\frac{\bar{b}_i}{\bar{a}_{ik}} \mid \bar{a}_{ik} > 0 \right\} \leq \Delta b_k \leq \min_i \left\{ -\frac{\bar{b}_i}{\bar{a}_{ik}} \mid \bar{a}_{ik} < 0 \right\}.$$

10.5.3 工艺条件(技术系数) A 的变化分析

设 A 的 l 行 k 列元 a_{lk} 有改变 Δa_{lk} .

(1) 当 a_{lk} 所在的列向量为非基向量时, Δa_{lk} 不影响解的可行性, 只要对应的

检验数 $\sigma'_k = \sigma_k - C_B B^{-1} \Delta a_{ik} \leq 0$, 则可得 $C_B B^{-1} \Delta a_{ik} \geq \sigma_k$, 即 $C_B \beta_k \Delta a_{ik} \geq \sigma_k$, 其中 $B^{-1} = (\beta_1, \cdots, \beta_k, \cdots, \beta_m)$.

于是, 当 $C_B \beta_k > 0$ 时, $\Delta a_{ik} \geq \frac{\sigma_k}{C_B \beta_k}$; 当 $C_B \beta_k < 0$ 时, $\Delta a_{ik} \leq \frac{\sigma_k}{C_B \beta_k}$.

(2) 当 a_{ik} 所在的列为基变量时, 由于 Δa_{ik} 不仅影响解的可行性, 而且会影响解的最优性, 情况比较复杂, 对具体问题只能具体分析了.

10.6 南水北调水指标的分配问题

10.6.1 问题的提出

南水北调中线工程建成后, 预计 2010 年年调水量为 110 亿 m^3 , 主要用来解决京、津、冀、豫四省(市)的沿线 20 个大中城市的生活用水、工业用水和综合服务业的用水, 分配比例分别为 40%、38% 和 22%. 这样可以改善我国中部地区的生态环境和投资环境, 推动经济发展. 用水指标的分配总原则是: 改善区域的缺水状况、提高城市的生活水平、促进经济发展、提高用水效益、改善城市环境. 根据 2000 年的统计数据, 各城市的人口数量差异大, 基本状况和经济情况也不相同. 各城市现有的生活、工业和综合服务业的用水情况不同, 缺水程度也不同(如表 10-4).

表 10-4 2000 年的基本情况

| 序号 | 城市名称 | 城市人口 | | 工业产值 | | 综合服务业总产值 | | 人均生活用水量 (升/日) | 万元综合服务业用水量 (m^3) | 万元工业增加值用水量 (m^3) |
|----|------|------------|---------------|-------------|------------|----------|-------------|------------------|--------------------------------|--------------------------------|
| | | 总数 (万人) | 年自然增长率 (%) | 增加值 (亿元) | 增长率 (%) | 人均产值(万元) | 年增长率 (%) | | | |
| 1 | 北京 | 1 285 | 2.04 | 737 | 11.1 | 1.16 | 13.2 | 354 | 160 | 143 |
| 2 | 天津 | 682 | 3.03 | 739 | 11.7 | 0.83 | 12.2 | 209 | 140 | 72 |
| 3 | 廊坊 | 56 | 9.15 | 193 | 10.0 | 0.30 | 10.0 | 245 | 180 | 102 |
| 4 | 保定 | 87 | 5.9 | 268 | 12.5 | 0.23 | 12.1 | 325 | 360 | 96 |
| 5 | 沧州 | 46 | 5.87 | 480 | 9.8 | 0.22 | 8.6 | 185 | 315 | 110 |
| 6 | 衡水 | 78 | 6.12 | 256 | 7.6 | 0.2 | 7.6 | 178 | 318 | 120 |
| 7 | 石家庄 | 218 | 5.41 | 464 | 10.2 | 0.44 | 11.8 | 267 | 235 | 86 |

续表

| 序号 | 城市名称 | 城市人口 | | 工业产值 | | 综合服务业总产值 | | 人均生活用水量 (升/日) | 万元综合服务业用水量 (m ³) | 万元工业增加值用水量 (m ³) |
|-------|------|------------|---------------|-------------|-------------|--------------|-------------|------------------|---------------------------------|---------------------------------|
| | | 总数 (万人) | 年自然增长率 (‰) | 增加值 (亿元) | 年增长率 (%) | 人均产值 (万元) | 年增长率 (%) | | | |
| 8 | 邢台 | 52 | 4.5 | 189 | 10.9 | 0.15 | 12.0 | 165 | 315 | 131 |
| 9 | 邯郸 | 81 | 3.69 | 721 | 10.0 | 0.22 | 11.4 | 230 | 320 | 126 |
| 10 | 安阳 | 83 | 6.61 | 110 | 6.9 | 0.16 | 11.7 | 320 | 310 | 186 |
| 11 | 鹤壁 | 42 | 8.0 | 36 | 9.6 | 0.18 | 12.3 | 220 | 320 | 210 |
| 12 | 濮阳 | 41 | 6.1 | 97 | 8.7 | 0.14 | 13.5 | 174 | 352 | 170 |
| 13 | 焦作 | 72 | 6.01 | 104 | 10.3 | 0.22 | 8.9 | 160 | 280 | 205 |
| 14 | 新乡 | 128 | 6.92 | 67 | 8.0 | 0.18 | 9.4 | 250 | 310 | 180 |
| 15 | 郑州 | 220 | 5.12 | 310 | 12.9 | 0.53 | 10.2 | 164 | 220 | 88 |
| 16 | 许昌 | 78 | 6.56 | 72 | 11.1 | 0.17 | 9.2 | 180 | 320 | 210 |
| 17 | 平顶山 | 90 | 6.61 | 114 | 8.8 | 0.18 | 8.4 | 155 | 310 | 189 |
| 18 | 周口 | 32 | 6.44 | 106 | 10.0 | 0.12 | 11.3 | 165 | 340 | 210 |
| 19 | 漯河 | 58 | 4.6 | 83 | 9.0 | 0.15 | 10.3 | 148 | 280 | 200 |
| 20 | 南阳 | 121 | 5.9 | 211 | 10.4 | 0.13 | 8.8 | 202 | 320 | 180 |
| 全国平均值 | | — | 10.7 | — | 9.9 | 0.23 | 7.8 | 219 | 610 | 288 |

要研究的问题是：

(1) 请你综合考虑各种情况，给出 2010 年每个城市的调水分配指标，使得各城市的总用水情况尽量均衡；

(2) 由于各城市的基本状况和自然条件不同，对相同的供水量所产生的经济效益不同，请从经济效益的角度，给出调水指标的分配方案。但是，要注意到，每个城市的工业和综合服务业的发展受产业规模的限制，不可能在短时间内无限制的增长。

10.6.2 模型的假设与符号说明

1. 模型的假设

(1) 原有供水量基本保持不变；

(2) 人口自然增长率、工业增加值年增长率和综合服务业人均产值年增长率基本保持不变；

(3) 所给工业增加值年增长率、综合服务业人均产值年增长率是在 2000 年的供水条件下的值,即投入的水量按当年的万元用水量产生经济效益。

2. 符号说明

$P_i^{(0)}, P_i^{(1)}$ 分别表示第 i 个城市的 2000 年, 2010 年的人口总数; Pr_i 表示第 i 个城市人口自然增长率; $I_i^{(0)}, I_i^{(1)}$ 分别表示第 i 个城市 2000 年, 2010 年的工业增加值; Ir_i 表示第 i 个城市工业年增长率; $S_i^{(0)}, S_i^{(1)}$ 分别表示第 i 个城市 2000 年, 2010 年的综合服务业人均产值; Sr_i 表示第 i 个城市综合服务业年增长率; W_i 表示第 i 个城市 2010 年的分配用水量; $X_i^{(0)}, X_i^{(1)}$ 分别表示第 i 个城市 2000 年, 2010 年的人均生活用水量; $Y_i^{(0)}, Y_i^{(1)}$ 分别表示第 i 个城市 2000 年, 2010 年的万元工业增加值用水量; $Z_i^{(0)}, Z_i^{(1)}$ 分别表示第 i 个城市的 2000 年, 2010 年的万元综合服务业用水量; x_i 表示分配第 i 个城市的生活用水总量指标; y_i 表示分配第 i 个城市的工业用水总量指标; z_i 表示分配第 i 个城市的综合服务业用水总量指标; a, b, c 分别表示生活用水, 工业用水, 综合服务业用水的分配比例。

10.6.3 问题的分析

对于问题(1), 要求各城市的总用水情况尽量“均衡”, 而各城市现有的三项用水指标各不相同, 因此, 我们把“均衡”定义为各城市新增加供水量与原有供水量的比例相等。

对于问题(2), 需要给出一个供水指标的优化分配方案, 这可以通过建立线性规划模型来实现。首先, 根据各城市的实际数据, 可以计算出 2010 年的每个城市的工业和综合服务业万元产值用水量。工业万元增加值用水量为

$$Y_i^{(1)} = \frac{Y_i^{(0)} \cdot I_i^{(0)}}{I_i^{(0)} \cdot (1 + Ir_i)^{10}} = \frac{Y_i^{(0)}}{(1 + Ir_i)^{10}} \quad (i = 1, 2, \dots, 20),$$

即 2010 年工业万元增加值用水量等于 2000 年的工业用水总量除以 2010 年的预计工业增加值(不考虑调水的情况下)。那么, 调水后用于工业产生的经济效益等于工业的调水总量除以万元增加值用水量, 即 $GI_i = \frac{y_i}{Y_i^{(1)}} \quad (i = 1, 2, \dots,$

20)。对于综合服务业可得到类似结果:

$$GS_i = \frac{z_i}{Z_i^{(1)}} \quad (i = 1, 2, \dots, 20),$$

$$\text{其中} \quad Z_i^{(1)} = \frac{Z_i^{(0)} \cdot S_i^{(0)}}{S_i^{(0)} \cdot (1 + Sr_i)^{10} (1 + Pr_i)^{10}} = \frac{Z_i^{(0)}}{(1 + Sr_i)^{10} (1 + Pr_i)^{10}} \quad (i = 1, 2, \dots, 20).$$

然后, 以调水量产生的工业和综合服务业效益总值为最大化目标, 调水总量为约束, 建立线性规划模型。

注：工业和综合服务业的发展受产业规模的限制，不可能在短时间内无限制的增长。因此，要适当考虑各城市的工业和综合服务业的均衡发展，即调水量不应过分集中。参照问题(1)，引入各城市调水量与原有供水量的比值（工业和综合服务业）作为衡量指标，并限制每个城市的用于工业和综合服务业的调水指标都不能低于平均值（问题(1)的计算结果）的50%。将这一约束加入到线性规划模型中，即可求得分配方案。

10.6.4 模型的建立与求解

问题(1)：

首先确定2010年受益城市的生活用水、工业用水、综合服务业平均增长的比例分别为 $\lambda_1, \lambda_2, \lambda_3$ 。由假设可有

$$\text{生活用水: } \sum_{i=1}^{20} X_i^{(0)} \cdot P_i^{(0)} + 110 \times 10^8 \times 0.4 = \sum_{i=1}^{20} X_i^{(0)} \cdot P_i^{(0)} \cdot (1 + \lambda_1),$$

$$\text{工业用水: } \sum_{i=1}^{20} Y_i^{(0)} \cdot I_i^{(0)} + 110 \times 10^8 \times 0.38 = \sum_{i=1}^{20} Y_i^{(0)} \cdot I_i^{(0)} \cdot (1 + \lambda_2),$$

$$\text{综合服务业用水: } \sum_{i=1}^{20} Z_i^{(0)} \cdot S_i^{(0)} + 110 \times 10^8 \times 0.22 = \sum_{i=1}^{20} Z_i^{(0)} \cdot S_i^{(0)} \cdot (1 + \lambda_3),$$

则求解得

$$\lambda_1 = \frac{110 \times 10^8 \times 0.4}{\sum_{i=1}^{20} X_i^{(0)} \cdot P_i^{(0)}}, \quad \lambda_2 = \frac{110 \times 10^8 \times 0.38}{\sum_{i=1}^{20} Y_i^{(0)} \cdot I_i^{(0)}}, \quad \lambda_3 = \frac{110 \times 10^8 \times 0.22}{\sum_{i=1}^{20} Z_i^{(0)} \cdot S_i^{(0)}}.$$

经计算可得 $\lambda_1 = 1.3011, \lambda_2 = 0.63169, \lambda_3 = 0.56856$ 。然后，按照“均衡”原则可以计算出各城市的生活用水、工业用水、综合服务业用水的分配指标，即

$$x_i = X_i^{(0)} \cdot P_i^{(0)} \cdot \lambda_1 \quad (i = 1, 2, \dots, 20),$$

$$y_i = Y_i^{(0)} \cdot I_i^{(0)} \cdot \lambda_2 \quad (i = 1, 2, \dots, 20),$$

$$z_i = Z_i^{(0)} \cdot S_i^{(0)} \cdot \lambda_3 \quad (i = 1, 2, \dots, 20),$$

故分配用水总量为 $W_i = x_i + y_i + z_i \quad (i = 1, 2, \dots, 20)$ ，计算结果如表10-5。

表 10-5 问题(1)调水分配方案

| 城市序号 水指标 (亿 m ³) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 生活用水 x_i | 21.15 | 6.769 | 0.652 | 1.343 | 0.404 | 0.659 | 2.764 | 0.407 | 0.885 | 1.261 |
| 工业用水 y_i | 6.657 | 3.361 | 1.244 | 1.625 | 3.335 | 1.941 | 2.521 | 1.564 | 5.739 | 1.292 |
| 服务业用水 z_i | 13.28 | 4.506 | 0.172 | 0.41 | 0.181 | 0.282 | 1.282 | 0.14 | 0.324 | 0.234 |
| 总用水量 W_i | 41.44 | 14.64 | 2.068 | 3.379 | 3.92 | 2.882 | 6.567 | 2.111 | 6.948 | 2.787 |

续表

| 城市序号 水指标 (亿 m ³) | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|------------------------------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 生活用水 x_i | 0.439 | 0.339 | 0.547 | 1.52 | 1.714 | 0.667 | 0.662 | 0.251 | 0.408 | 1.161 |
| 工业用水 y_i | 0.478 | 1.042 | 1.347 | 0.762 | 1.723 | 0.955 | 1.361 | 1.406 | 1.046 | 2.399 |
| 服务业用水 y_i | 0.138 | 0.115 | 0.252 | 0.406 | 1.459 | 0.241 | 0.286 | 0.074 | 0.139 | 0.286 |
| 总用水量 w_i | 1.055 | 1.494 | 2.146 | 2.688 | 4.896 | 1.863 | 2.314 | 1.731 | 1.593 | 3.846 |

问题(2):

根据上面对问题的分析,生活用水指标与问题(1)相同,只需讨论工业和综合服务业的调水指标.由于各城市的基本状况和自然条件的差异,对相同的供水量所产生的经济效益不同,从经济效益的角度力求对调水指标有最高的经济效益.同时注意到,每个城市的工业和综合服务业的发展受产业规模的限制,不可能在短时间内无限制的增长.由此可知,对每个城市的调水指标都应有上限和下限的约束.于是,2010 年最优的调水分配指标应满足下面的线性规划模型.

目标函数为总的经济效益最大,即

$$\max G = \sum_{i=1}^{20} GI_i + \sum_{i=1}^{20} GS_i = \sum_{i=1}^{20} \frac{y_i}{Y_i^{(1)}} + \sum_{i=1}^{20} \frac{z_i}{Z_i^{(1)}},$$

其约束条件为

$$\begin{cases} \sum_{i=1}^{20} y_i = 0.38 \times 110 \times 10^8, \\ \sum_{i=1}^{20} z_i = 0.22 \times 110 \times 10^8, \\ 0.5\lambda_2 \leq \frac{y_i}{Y_i^{(0)} \cdot I_i^{(0)}} \leq 1.5\lambda_2, \\ 0.5\lambda_3 \leq \frac{z_i}{Z_i^{(0)} \cdot S_i^{(0)}} \leq 1.5\lambda_3, \\ y_i, z_i \geq 0 \quad (i = 1, 2, \dots, 20), \end{cases}$$

其中 $Y_i^{(1)} = \frac{Y_i^{(0)}}{(1 + Ir_i)^{10}}, Z_i^{(1)} = \frac{Z_i^{(0)}}{(1 + Sr_i)^{10}(1 + Pr_i)^{10}} \quad (i = 1, 2, \dots, 20)$. 这是一个较复杂的线性规划模型,使用 MATLAB 的 linprog 函数求解可得最优值为 $G = 7.375\ 80 \times 10^{18}$. 具体的调水分配方案如表 10-6.

表 10-6 问题(2)的调水分配方案

| 城市序号 水指标 (亿立方米) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------------|--------|--------|-------|-------|-------|-------|-------|-------|-------|-------|
| 生活用水量 x_i | 21.15 | 6.769 | 0.652 | 1.343 | 0.404 | 0.659 | 2.764 | 0.407 | 0.885 | 1.261 |
| 工业用水量 y_i | 9.986 | 5.042 | 1.865 | 2.438 | 1.668 | 0.97 | 3.781 | 2.346 | 2.869 | 0.646 |
| 服务业用水量 y_i | 18.737 | 2.253 | 0.086 | 0.205 | 0.091 | 0.141 | 0.641 | 0.07 | 0.162 | 0.117 |
| 总用水量 w_i | 49.873 | 14.064 | 2.603 | 3.986 | 2.163 | 1.77 | 7.186 | 2.823 | 3.916 | 2.024 |
| 城市序号 水指标 (亿立方米) | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| 生活用水量 x_i | 0.439 | 0.339 | 0.547 | 1.52 | 1.714 | 0.667 | 0.662 | 0.251 | 0.408 | 1.161 |
| 工业用水量 y_i | 0.239 | 0.521 | 0.673 | 0.381 | 2.585 | 1.433 | 0.680 | 0.703 | 0.524 | 2.449 |
| 服务业用水量 y_i | 0.069 | 0.057 | 0.126 | 0.203 | 0.729 | 0.121 | 0.143 | 0.037 | 0.069 | 0.143 |
| 总用水量 w_i | 0.747 | 0.917 | 1.346 | 2.104 | 5.028 | 2.221 | 1.485 | 0.991 | 1.001 | 3.753 |

10.7 参考案例与参考文献

1. 参考案例

- (1) 豆油厂的合理布局问题——文献[3]:74—77
- (2) 森林资源的管理问题——文献[4]:126—129
- (3) 投资组合问题——文献[5]:175—177
- (4) 基金使用计划问题——文献[7]:77—84
- (5) 奶制品加工计划的设计问题——文献[8]:40—42

2. 参考文献

- [1] 叶其孝. 大学生数学建模竞赛辅导教材(二). 长沙:湖南教育出版社,1997
- [2] 姜启源. 数学模型. 第三版. 北京:高等教育出版社,2003
- [3] 寿纪麟. 数学建模方法与范例. 西安:西安交通大学出版社,1993
- [4] 《数学模型》编写组. 数学模型. 广州:华南理工大学出版社,2003
- [5] 何坚勇. 运筹学基础. 北京:清华大学出版社,1999
- [6] 杨超学. 运筹学. 北京:科学出版社,2004
- [7] 刘承平. 数学建模方法. 北京:高等教育出版社,2002
- [8] 施光燕等. 最优化方法. 北京:高等教育出版社,1999

- [9] 《运筹学》教材编写组, 运筹学. 修订版. 北京: 清华大学出版社, 1990
- [10] 韩中庚, 姜启源. 奶制品加工计划的设计模型. 信息工程大学学报, 2002, 3(1): 40—42

第十一章 整数规划方法

11.1 整数规划模型

如果一个数学规划的某些决策变量或全部决策变量要求必须取整数,则这样的问题称为**整数规划问题**,其模型称为**整数规划模型**.

如果整数规划的目标函数和约束都是线性的,则称此问题为**整数线性规划问题**.在这里我们只就整数线性规划问题进行讨论,整数线性规划的一般模型为

$$\begin{aligned} \max(\min) z &= \sum_{j=1}^n c_j x_j, \\ \text{s. t. } \begin{cases} \sum_{j=1}^n a_{ij} x_j \leq (=, \geq) b_i & (i = 1, 2, \dots, m), \\ x_j \geq 0, x_j \text{ 为整数} & (j = 1, 2, \dots, n). \end{cases} \end{aligned} \quad (11.1)$$

对于实际中的某些整数规划问题,我们有时候可以想到先略去整数约束的条件,即视为一个线性规划问题,利用单纯形法求解,然后对其最优解进行取整处理.实际上,这样得到的解未必是原整数规划问题的最优解,因此,这种方法是不可取的,但可借鉴这种思想.

整数规划求解方法总的基本思想是:松弛问题(11.1)中的约束条件(譬如去掉整数约束条件),使构成易于求解的新问题——松弛问题(A),如果这个问题(A)的最优解是原问题(11.1)的可行解,则就是原问题(11.1)的最优解;否则,在保证不改变松弛问题(A)的可行性的条件下,修正松弛问题(A)的可行域(增加新的约束),变成新的问题(B),再求问题(B)的解,重复这一过程直到修正问题的最优解在原问题(11.1)的可行域内为止,即得到了原问题的最优解.

注:如果每个松弛问题的最优解不是原问题的可行解,则这个解对应的目标函数值 \bar{z} 一定是原问题最优值 z^* 的上界(最大化问题),即 $z^* \leq \bar{z}$;或下界(最小化问题),即 $z^* \geq \bar{z}$.

11.2 整数规划的分枝定界法

11.2.1 分枝定界法的基本思想

将原问题(11.1)中整数约束去掉变为问题(A),求出问题(A)的最优解,如果它不是原问题(11.1)的可行解,则通过附加线性不等式约束(整型),将问题(A)分枝变为若干子问题(B_i))($i=1,2,\dots,I$),即对每一个非整数变量附加两个互相排斥(不交叉)的整型约束,即可得到两个子问题,继续求解定界,重复这一过程,直到得到最优解为止.

例 11.1 用分枝定界法求解下列整数规划问题

$$\begin{aligned} \max z &= 40x_1 + 90x_2, \\ (A) \quad &\begin{cases} 9x_1 + 7x_2 \leq 56, \\ 7x_1 + 20x_2 \leq 70, \\ x_1, x_2 \geq 0, \text{且为整数.} \end{cases} \end{aligned}$$

解 (1) 去掉 x_1, x_2 为整数的约束,变为问题(A_0),求解可得(A_0)的最优解为 $x_1 = 4.81, x_2 = 1.82, z_0 = 356 = \bar{z}$,令 $\underline{z} = 0$,且 $0 = \underline{z} \leq z^* \leq \bar{z} = 356$. 显然不是(A)的可行解.

(2) 分枝:对 $x_1 = 4.81$ 附加约束条件 $x_1 \leq 4, x_1 \geq 5$,将(A)分为两个子问题(B_1)和(B_2):

$$\begin{aligned} \max z &= 40x_1 + 90x_2, & \max z &= 40x_1 + 90x_2, \\ (B_1) \quad &\begin{cases} 9x_1 + 7x_2 \leq 56, \\ 7x_1 + 20x_2 \leq 70, \\ 0 \leq x_1 \leq 4, x_2 \geq 0; \end{cases} & (B_2) \quad &\begin{cases} 9x_1 + 7x_2 \leq 56, \\ 7x_1 + 20x_2 \leq 70, \\ x_1 \geq 5, x_2 \geq 0. \end{cases} \end{aligned}$$

求解可得(B_1): $x_1 = 4, x_2 = 2.1, z_1 = 349$; (B_2): $x_1 = 5, x_2 = 1.57, z_2 = 341$. 仍不是问题(A)的可行解,取 $\bar{z} = 341, \underline{z} = 0$.

(3) 对问题(B_1), (B_2)分枝:

对问题(B_1): $x_2 = 2.1$ 附加约束条件 $x_2 \leq 2, x_2 \geq 3$ 分为问题(C_1), (C_2).

对问题(B_2): $x_2 = 1.57$ 附加约束条件 $x_2 \leq 1, x_2 \geq 2$ 分为问题(C_3), (C_4).

$$\begin{aligned} \max z &= 40x_1 + 90x_2, & \max z &= 40x_1 + 90x_2, \\ (C_1) \quad &\begin{cases} 9x_1 + 7x_2 \leq 56, \\ 7x_1 + 20x_2 \leq 70, \\ 0 \leq x_1 \leq 4, 0 \leq x_2 \leq 2; \end{cases} & (C_2) \quad &\begin{cases} 9x_1 + 7x_2 \leq 56, \\ 7x_1 + 20x_2 \leq 70, \\ 0 \leq x_1 \leq 4, x_2 \geq 3; \end{cases} \end{aligned}$$

$$\begin{aligned} \max z &= 40x_1 + 90x_2, & \max z &= 40x_1 + 90x_2, \\ (C_3) \quad \begin{cases} 9x_1 + 7x_2 \leq 56, \\ 7x_1 + 20x_2 \leq 70, \\ x_1 \geq 5, 0 \leq x_2 \leq 1; \end{cases} & (C_4) \quad \begin{cases} 9x_1 + 7x_2 \leq 56, \\ 7x_1 + 20x_2 \leq 70, \\ x_1 \geq 5, x_2 \geq 2. \end{cases} \end{aligned}$$

分别求解可得

$$(C_1): x_1 = 4, x_2 = 2, z_3 = 340; \quad (C_2): x_1 = 1.42, x_2 = 3, z_4 = 327;$$

$$(C_3): x_1 = 5.44, x_2 = 1, z_5 = 308; \quad (C_4): \text{无可行解.}$$

因为 (C_1) 的解为 (A) 的可行解,故取 $\underline{z} = 340, \bar{z} = 341$,注意到 $z_4, z_5 \leq \underline{z} = 340$,则对 (C_2) 和 (C_3) 无需再分枝,将其剪掉.由 $\underline{z} = 340 \leq z^* \leq \bar{z} = 341$ 可知,已得到了问题 (A) 的最优解 $x_1^* = 4, x_2^* = 2, z^* = 340$.

11.2.2 分枝定界法的一般步骤

(1) 将原整数规划问题(11.1)去掉所有的整数约束变为线性规划问题 (A) ,用线性规划的方法求解问题 (A) ,则有下列情况:

- 1) 问题 (A) 无可行解,则原问题(11.1)也无可行解,停止计算;
- 2) 问题 (A) 有最优解 X^* ,并是原问题(11.1)的可行解,则此解就是 (A) 的最优解,计算结束;
- 3) 问题 (A) 有最优解 X^* ,但不是原问题(11.1)的可行解,转下一步.

(2) 将 X^* 代入目标函数,其值记为 \bar{z} ,并用观察法找出原问题(11.1)的一个可行解(整数解,不妨可取 $x_j = 0 \ (j = 1, 2, \dots, n)$),求得目标函数值(下界值),记为 \underline{z} ,则原问题(11.1)的最优值记为 z^* ,即有 $\underline{z} \leq z^* \leq \bar{z}$,转下一步;

(3) 分枝:在问题 (A) 的最优解中任选一个不满足整数约束的变量 $x_j = b_j$ (非整数),附加两个整数不等式约束: $x_j \leq [b_j]$ 和 $x_j \geq [b_j] + 1$,分别加入到问题 (A) 中,构成两个新的子问题 (B_1) 和 (B_2) ,仍不考虑整数约束,求问题 (B_1) 和 (B_2) 的解.

定界:对每一个子问题的求解结果,找出最优值的最大者为新的上界 \bar{z} ,从所有符合整数约束条件的分枝中找出目标函数值最大的一个为新的下界 \underline{z} .

(4) 比较与剪枝:将各分枝问题的最优值同 \underline{z} 比较,如果其值小于 \underline{z} ,则这个分枝可以剪掉,以后不再考虑.如果其值大于 \underline{z} ,且又不是原问题(11.1)的可行解,则继续分枝,返回(3),直到最后得到最优解使 $z^* = \underline{z}$,即 $x_j^* \ (j = 1, 2, \dots, n)$ 为原问题(11.1)的最优解.

11.3 整数规划的割平面法

11.3.1 割平面法的基本思想

首先把原整数规划问题(11.1)去掉整数约束条件变为线性规划问题(A), 然后引入线性约束条件(称为 **Gomory** 约束, 几何术语称为割平面)使问题(A)的可行域逐步缩小(即切割掉一部分), 每次切割掉的是问题非整数解的一部分, 不切掉任何整数解, 直到最后使得目标函数达到最优的整数解(点)成为可行域的一个顶点为止, 这也就是原问题(11.1)的最优解. 即利用线性规划的求解方法逐步缩小可行域, 最后找到整数规划问题的最优解.

例 11.2 用割平面法求解下列整数规划问题

$$\begin{aligned} \max z &= x_1 + x_2, \\ \begin{cases} -x_1 + x_2 \leq 1, \\ 3x_1 + x_2 \leq 4, \\ x_1, x_2 \geq 0, \text{且为整数.} \end{cases} \end{aligned}$$

将其化为标准型

$$\begin{aligned} \max z &= x_1 + x_2, \\ (A) \quad \begin{cases} -x_1 + x_2 + x_3 = 1, \\ 3x_1 + x_2 + x_4 = 4, \\ x_1, x_2, x_3, x_4 \geq 0, \text{且为整数.} \end{cases} \end{aligned}$$

去掉整数约束条件, 用单纯形法求解相应的线性规划问题, 可得非整数的最优解

为 $x_1 = \frac{3}{4}, x_2 = \frac{7}{4}, x_3 = x_4 = 0, z = \frac{5}{2}$, 由变量之间的关系可以得到关系式

$$\begin{cases} x_1 - \frac{1}{4}x_3 + \frac{1}{4}x_4 = \frac{3}{4}, \\ x_2 + \frac{3}{4}x_3 + \frac{1}{4}x_4 = \frac{7}{4} = 1 + \frac{3}{4}, \end{cases}$$

将整数与真分式分开有

$$\begin{cases} x_1 - x_3 = \frac{3}{4} - \left(\frac{3}{4}x_3 + \frac{1}{4}x_4 \right), \\ x_2 - 1 = \frac{3}{4} - \left(\frac{3}{4}x_3 + \frac{1}{4}x_4 \right). \end{cases}$$

因为 $x_1, x_2, x_3, x_4 \geq 0$, 且为整数, 即左端为整数, 右端也必为整数, 所以 $\frac{3}{4} -$

$\left(\frac{3}{4}x_3 + \frac{1}{4}x_4 \right) \leq 0$, 即 $-3x_3 - x_4 \leq -3$ 为所求的切割方程. 将其加到问题的约束

条件之中,构成新的问题(B),并引入松弛变量 x_5 ,则有 $-3x_3 - x_4 + x_5 = -3$,即求问题

$$(B) \begin{cases} \max z = x_1 + x_2, \\ -x_1 + x_2 + x_3 = 1, \\ 3x_1 + x_2 + x_4 = 4, \\ -3x_3 - x_4 + x_5 = -3, \\ x_1, x_2, x_3, x_4, x_5 \geq 0. \end{cases}$$

用单纯形法求解,则可得到问题(B)的最优解为 $x_1 = x_2 = x_3 = 1$,显然是原问题的可行解,即为所求整数规划问题的最优解,其最优值为 $z = 2$ 。

11.3.2 割平面法的计算步骤

设原问题(11.1)中有 n 个决策变量, m 个松弛变量,共 $n+m$ 个变量,略去整数约束求解线性规划问题,将最终的求解结果放入如表11-1(称为单纯形表)。其中 x_i ($i=1,2,\dots,m$)表示基变量, y_j ($j=1,2,\dots,n$)表示非基变量,则具体计算步骤如下:

(1) 在最优解中任取一个具有分数值的基变量,不妨就是 x_i ($1 \leq i \leq m$),由表11-1可以得到关系 $x_i + \sum_{j=1}^n \bar{a}_{ij}y_j = \bar{b}_i$,即

$$x_i = \bar{b}_i - \sum_{j=1}^n \bar{a}_{ij}y_j \quad (1 \leq i \leq m); \quad (11.2)$$

表 11-1 单纯形表

| 基 | x_1 | ... | x_i | ... | x_m | y_1 | ... | y_j | ... | y_n | b |
|------------|----------|-----|----------|-----|----------|----------------|-----|----------------|-----|----------------|-------------|
| x_1 | 1 | ... | 0 | ... | 0 | \bar{a}_{11} | ... | \bar{a}_{1j} | ... | \bar{a}_{1n} | \bar{b}_1 |
| \vdots | \vdots | | \vdots | | \vdots | \vdots | | \vdots | | \vdots | \vdots |
| x_i | 0 | ... | 1 | ... | 0 | \bar{a}_{i1} | ... | \bar{a}_{ij} | ... | \bar{a}_{in} | \bar{b}_i |
| \vdots | \vdots | | \vdots | | \vdots | \vdots | | \vdots | | \vdots | \vdots |
| x_m | 0 | ... | 0 | ... | 1 | \bar{a}_{m1} | ... | \bar{a}_{mj} | ... | \bar{a}_{mn} | \bar{b}_m |
| σ_j | 0 | ... | 0 | ... | 0 | σ_1 | ... | σ_j | ... | σ_n | $-z$ |

(2) 将 \bar{b}_i 和 \bar{a}_{ij} ($1 \leq i \leq m; j=1,2,\dots,n$) (为假分数)分为整数部分和非负的真分数,即

$$\bar{b}_i = N_i + f_i, \quad \bar{a}_{ij} = N_{ij} + f_{ij} \quad (j=1,2,\dots,n),$$

其中 N_i 和 N_{ij} 表示整数,而 f_i ($0 < f_i < 1$)和 f_{ij} ($0 \leq f_{ij} < 1$)表示真分数,代入(11.2)式,并将整数放在一边,分数放在一边,即

$$x_i + \sum_{j=1}^n N_{ij} y_j - N_j = f_i - \sum_{j=1}^n f_{ij} y_j \quad (1 \leq i \leq m); \quad (11.3)$$

(3) 要使 x_i 和 y_j 都为整数, (11.3) 式的左端必为整数, 右端也是整数, 而且由 $f_{ij} \geq 0$, y_j 是非负整数, 故此 $\sum_{j=1}^n f_{ij} y_j \geq 0$. 又因 $f_i > 0$ 是真分数, 于是有 $f_i -$

$\sum_{j=1}^n f_{ij} y_j \leq f_i < 1$, 则必有

$$f_i - \sum_{j=1}^n f_{ij} y_j \leq 0 \quad (1 \leq i \leq m), \quad (11.4)$$

这就是所要求的一个切割方程 (Gomory 约束条件);

(4) 对 (11.4) 式引入一个松弛变量 S_i , 则 (11.4) 式变为

$$S_i - \sum_{j=1}^n f_{ij} y_j = -f_i \quad (1 \leq i \leq m),$$

将其代入原问题中去, 求解新的线性规划问题;

(5) 应用对偶单纯形法求解, 如果所求最优解为原问题的可行解 (整数解), 则就是原问题的最优解, 计算结束. 否则继续构造 Gomory 约束条件, 直到其最优解为整数解停止.

说明: 在这里为什么要用对偶单纯形法求解呢? 主要是在 Gomory 方程 $S_i - \sum_{j=1}^n f_{ij} y_j = -f_i$ 中, 当非基变量 $y_j = 0$ 时, 条件变为 $S_i = -f_i$ 为负数, 即为不可行的, 通常的单纯形法无法求解, 而用对偶单纯形法从不可行到可行, 即可得到最优解.

11.4 0-1 整数规划

11.4.1 0-1 整数规划的模型

如果整数线性规划问题的所有决策变量 x_i 仅限于取 0 或 1 两个数值, 则称此问题为 **0-1 线性整数规划**, 简称为 **0-1 规划**. 变量 x_i 称为 **0-1 变量**, 或 **二进制变量**, 其变量取值的约束可变为 $x_i = 0$ 或 1, 等价于 $x_i \leq 1$ 和 $x_i \geq 0$ 且为整数. 于是 0-1 规划的一般模型为

$$\begin{aligned} \max(\min) z &= \sum_{j=1}^n c_j x_j, \\ \text{s. t. } \begin{cases} \sum_{j=1}^n a_{ij} x_j \leq (=, \geq) b_i & (i = 1, 2, \dots, m), \\ x_j = 0 \text{ 或 } 1 & (j = 1, 2, \dots, n). \end{cases} \end{aligned}$$

例 11.3 背包问题(或载货问题)

一个旅行者要在背包里装一些最有用的东西,但限制最多只能带 b kg 物品,每件物品只能是整件携带,对每件物品都规定了一定的“使用价值”(有用的程度). 如果共有 n 件物品,第 j 件物品重 a_j kg,其价值为 c_j ,问题是:在携带的物品总重量不超过 b kg 的条件下,携带哪些物品可使总价值最大?

设决策变量 $x_j: x_j = \begin{cases} 1, & \text{当携带第 } j \text{ 种物品时,} \\ 0, & \text{当不携带第 } j \text{ 种物品时,} \end{cases}$ 则问题的模型为

$$\begin{aligned} \max z &= \sum_{j=1}^n c_j x_j, \\ \text{s. t. } &\begin{cases} \sum_{j=1}^n a_j x_j \leq b, \\ x_j = 0 \text{ 或 } 1, \end{cases} \end{aligned}$$

即为一个 0-1 规划模型.

例 11.4 指派(或分配)问题

在生产管理上,为了完成某项任务,总是希望把有关人员最合理地分派,以发挥其最大工作效率,创造最大的价值.

例如: 设某单位有 4 个人,每个人都有能力去完成 4 项科研任务中的任一项,由于 4 个人的能力和经验不同,所需完成各项任务的时间如表 11-2 所示. 问分配何人去完成何项任务使完成所有任务的总时间最少?

表 11-2 4 个人完成 4 项任务的时间

| 项目 人员 | A | B | C | D |
|----------|----|----|----|----|
| 甲 | 2 | 15 | 13 | 4 |
| 乙 | 10 | 4 | 14 | 15 |
| 丙 | 9 | 14 | 16 | 13 |
| 丁 | 7 | 8 | 11 | 9 |

设决策变量 x_{ij} 表示第 i 个人去完成第 j 项任务,即

$$x_{ij} = \begin{cases} 1, & \text{当第 } i \text{ 个人去完成第 } j \text{ 项任务时,} \\ 0, & \text{当第 } i \text{ 个人不去完成第 } j \text{ 项任务时} \end{cases} \quad (1 \leq i, j \leq 4).$$

每个人去完成一项任务的约束为

$$\begin{cases} x_{11} + x_{12} + x_{13} + x_{14} = 1, \\ x_{21} + x_{22} + x_{23} + x_{24} = 1, \\ x_{31} + x_{32} + x_{33} + x_{34} = 1, \\ x_{41} + x_{42} + x_{43} + x_{44} = 1. \end{cases}$$

每一项任务必有一人去完成的约束为

$$\begin{cases} x_{11} + x_{21} + x_{31} + x_{41} = 1, \\ x_{12} + x_{22} + x_{32} + x_{42} = 1, \\ x_{13} + x_{23} + x_{33} + x_{43} = 1, \\ x_{14} + x_{24} + x_{34} + x_{44} = 1. \end{cases}$$

目标函数为完成任务的总时间:

$$\begin{aligned} \min z = & 2x_{11} + 15x_{12} + 13x_{13} + 4x_{14} + 10x_{21} + 4x_{22} + 14x_{23} + 15x_{24} \\ & + 9x_{31} + 14x_{32} + 16x_{33} + 13x_{34} + 7x_{41} + 8x_{42} + 11x_{43} + 9x_{44} \end{aligned}$$

记系数矩阵为 $C = (c_{ij}) = \begin{bmatrix} 2 & 15 & 13 & 4 \\ 10 & 4 & 14 & 15 \\ 9 & 14 & 16 & 13 \\ 7 & 8 & 11 & 9 \end{bmatrix}$, 称为效益矩阵, 或价值矩阵,

c_{ij} 表示第 i 个人去完成第 j 项任务时有关的效益 (时间、费用、价值等). 故该问题的为一个 0-1 规划模型:

$$\begin{aligned} \min z = & \sum_{i=1}^4 \sum_{j=1}^4 c_{ij} x_{ij}, \\ \text{s. t. } & \begin{cases} \sum_{j=1}^4 x_{ij} = 1, i = 1, 2, 3, 4, \\ \sum_{i=1}^4 x_{ij} = 1, j = 1, 2, 3, 4, \\ x_{ij} = 0 \text{ 或 } 1 \quad (i, j = 1, 2, 3, 4). \end{cases} \end{aligned}$$

一般的指派 (或分配) 问题:

设某单位有 n 项任务, 正好需要 n 个人去完成, 由于各项任务的性质和每人的专长不同, 如果分配每个人仅能完成一项任务, 应如何分派使完成 n 项任务的总效益 (或效率) 最高.

设该指派问题有相应的效益矩阵 $C = (c_{ij})_{n \times n}$, 其元素 c_{ij} 表示分配第 i 个人去完成第 j 项任务时的效益. 或者说: 以 c_{ij} 表示给定的第 i 单位资源分配用于第 j 项活动时的有关效益.

设问题的决策变量为 x_{ij} 是 0-1 变量, 即

$$x_{ij} = \begin{cases} 1, & \text{当分配第 } i \text{ 个单位资源用于第 } j \text{ 项活动时,} \\ 0, & \text{当不分配第 } i \text{ 个单位资源用于第 } j \text{ 项活动时} \end{cases} \quad (i, j = 1, 2, \dots, n).$$

其数学模型为

$$\min z = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij},$$

$$\text{s. t. } \begin{cases} \sum_{j=1}^n x_{ij} = 1, i = 1, 2, \dots, n, \\ \sum_{i=1}^n x_{ij} = 1, j = 1, 2, \dots, n, \\ x_{ij} = 0 \text{ 或 } 1 \quad (i, j = 1, 2, \dots, n). \end{cases} \quad (11.5)$$

11.4.2 0-1 规划的隐枚举法

显枚举法(又称为穷举法):主要是把问题的所有可能的组合情况(共 2^n 种)列举出来进行比较,找到所需要的最优解。

隐枚举法:主要是从实际出发,在问题所有可能的组合取值中利用“过滤条件”排除一些不可能是最优解的情况,只需考查其中一部分组合就可以得到问题的最优解。因此,隐枚举法又称为**部分枚举法**。

隐枚举法不需要更多的理论和知识,当问题的维数 n 不是特别大时,隐枚举法是求解0-1规划的一种有效方法,但对于过滤条件的确定,要根据实际问题的具体情况具体分析而定。

11.5 指派问题的匈牙利方法

“匈牙利方法”最早是由匈牙利数学家康尼格(D. Konig)用来求矩阵中零元素的个数的一种方法,由此他证明了“矩阵中独立零元素的最多个数等于能覆盖所有零元素的最少直线数”。1955年由库恩(W. W. Kuhn)在求解著名的指派问题时,引用了这一结论,并对具体算法做了改进,仍然称为“匈牙利方法”。

11.5.1 匈牙利方法的基本思想

由于每个问题都有一个相应的效益矩阵,可以通过初等变换修改效益矩阵的行或列的元素,使得在每一行或每一列中至少有一个零元素,直到在不同行、不同列中至少有一个零元素,从而得到与这些零元素相对应的一个完全分配方案,这个方案是原问题的一个最优分配方案。

定理 11.1 (指派问题的最优性) 如果问题(11.5)的效益矩阵 $C = (c_{ij})_{n \times n}$ 的第 i 行、第 j 列中的每个元素分别减去一个常数 a, b 变为矩阵 $D = (d_{ij})_{n \times n}$,则以新的矩阵 D 为效益矩阵和新的目标函数与原效益矩阵 C 和原目标函数求得的最优解相同,最优值只差一个常数。

证 只要证明新目标函数和原目标函数值相差一个常数。

事实上,因为 $d_{ij} = c_{ij} - a - b (1 \leq i, j \leq n)$,则新的目标函数为

$$\begin{aligned}
 z' &= \sum_{i=1}^n \sum_{j=1}^n d_{ij} x_{ij} \\
 &= \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij} - a \sum_{j=1}^n x_{ij} - b \sum_{i=1}^n x_{ij} \\
 &= \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij} - (a + b) = z - (a + b),
 \end{aligned}$$

故二者相差一个常数 $a + b$, 最优解相同.

11.5.2 匈牙利方法的基本步骤

根据指派问题的最优性, “若从效益矩阵 $C = (c_{ij})_{n \times n}$ 的一行(或列)各元素分别减去该行(列)的最小元素, 得到新矩阵 $D = (d_{ij})_{n \times n}$, 那么以 D 为效益矩阵所对应问题的最优解与原问题的最优解相同”. 此时求最优解的问题可转化为求效益矩阵的最大 1 元素组的问题.

下面给出一般的匈牙利方法的计算步骤:

第 1 步: 对效益矩阵进行变换, 使每行每列都出现有 0 元素.

(1) 从效益矩阵 C 中每一行减去该行的最小元素;

(2) 再在所得矩阵中每一列减去该列的最小元素, 所得矩阵记为 $D = (d_{ij})_{n \times n}$.

第 2 步: 将矩阵 D 中 0 元素置为 1 元素, 非零元置为 0 元素, 记此矩阵为 E .

第 3 步: 确定独立 1 元素组.

(1) 在矩阵 E 含有 1 元素的各行中选择 1 元素最少的行, 比较该行中各 1 元素所在的列中 1 元素的个数, 选择 1 元素的个数最少的一列的那个 1 元素;

(2) 将所选的 1 元素所在的行和列清 0;

(3) 重复第 2 步和第 3 步, 直到没有 1 元素为止, 即得到一个独立 1 元素组.

第 4 步: 判断是否为最大独立 1 元素组.

(1) 如果所得独立 1 元素组是原效益矩阵的最大独立 1 元素组(即 1 元素的个数等于矩阵的阶数), 则已得到最优解, 停止计算.

(2) 如果所得独立 1 元素组还不是原效益矩阵的最大独立 1 元素组, 那么利用寻找可扩路的方法对其进行扩张, 进行下一步.

第 5 步: 利用寻找可扩路方法确定最大独立 1 元素组.

(1) 做最少的直线覆盖矩阵 D 的所有 0 元素;

(2) 在没有被直线覆盖的部分找出最小元素, 在没有被直线覆盖的各行减去此最小元素, 在没被直线覆盖的各列加上此最小元素, 得到一个新的矩阵, 返回第 2 步.

说明:上面的算法是按最小化问题给出的,如果问题是最大化问题,即模型

(11.5)中的目标函数换为 $\max z = \sum_{i=1}^n \sum_{j=1}^n c_{ij} x_{ij}$. 此令 $M = \max_{i,j} (c_{ij})$ 和 $b_{ij} = M - c_{ij}$

≥ 0 ,则效率矩阵变为 $B = (b_{ij})_{n \times n}$. 于是考虑目标函数为 $\min z = \sum_{i=1}^n \sum_{j=1}^n b_{ij} x_{ij}$ 的问题,仍用上面的方法步骤求解所得最小解也就是对应原问题的最大解.

另外,按照匈牙利方法计算步骤,对于这一类问题的求解,可以用 MATLAB 编程实现.

11.6 招聘公务员问题^{①[8]}

11.6.1 问题的提出

我国公务员制度已实施多年,1993年10月1日颁布施行的《国家公务员暂行条例》规定:“国家行政机关录用担任主任科员以下的非领导职务的国家公务员,采用公开考试、严格考核的办法,按照德才兼备的标准择优录用”.目前,我国招聘公务员的程序一般分三步进行:公开考试(笔试)、面试考核、择优录取.

现有某市直属单位因工作需要,拟向社会公开招聘8名公务员,具体的招聘办法和程序如下:

(一)公开考试:凡是年龄不超过30周岁,大学专科以上学历,身体健康者均可报名参加考试,考试科目有:综合基础知识、专业知识和“行政职业能力测验”三个部分,每科满分为100分.根据考试总分的高低排序按1:2的比例(共16人)选择进入第二阶段的面试考核.

(二)面试考核:面试考核主要考核应聘人员的知识面、对问题的理解能力、应变能力、表达能力等综合素质.按照一定的标准,面试专家组对每个应聘人员的各个方面都给出一个等级评分,从高到低分成A/B/C/D四个等级,具体结果见表11-3所示.

(三)由招聘领导小组综合专家组的意见、笔初试成绩以及各用人部门需求确定录用名单,并分配到各用人部门.

该单位拟将录用的8名公务员安排到所属的7个部门,并且要求每个部门至少安排一名公务员.这7个部门按工作性质可分为四类:(1)行政管理、(2)技术管理、(3)行政执法、(4)公共事业.见表11-4所示.

招聘领导小组在确定录用名单的过程中,本着公平、公开的原则,同时考虑

① 本案例选自2004年中国大学生数学建模竞赛D题.

录用人员的合理分配和使用,有利于发挥个人的特长和能力.招聘领导小组将 7 个用人单位的基本情况(包括福利待遇、工作条件、劳动强度、晋升机会和学习深造机会等)和四类工作对聘用公务员的具体条件的希望达到的要求都向所有应聘人员公布(如表 11-4).每一位参加面试人员都可以申报两个自己的工作类别志愿(如表 11-3).请研究下列问题:

(1) 如果不考虑应聘人员的意愿,择优按需录用,试帮助招聘领导小组设计一种录用分配方案;

(2) 在考虑应聘人员意愿和用人单位的希望要求的情况下,请你帮助招聘领导小组设计一种分配方案;

(3) 你的方法对于一般情况,即 N 个应聘人员 M 个用人单位时,是否可行?

表 11-3 招聘公务员笔试成绩,专家组的面试评分及个人志愿

| 应聘人员 | 笔试成绩 | 申报类别志愿 | | 专家组对应聘者特长的等级评分 | | | |
|-------|------|--------|-----|----------------|--------|--------|------|
| | | | | 专业知识面 | 认识理解能力 | 灵活应变能力 | 表达能力 |
| 人员 1 | 290 | (2) | (3) | A | A | B | B |
| 人员 2 | 288 | (3) | (1) | A | B | A | C |
| 人员 3 | 288 | (1) | (2) | B | A | D | C |
| 人员 4 | 285 | (4) | (3) | A | B | B | B |
| 人员 5 | 283 | (3) | (2) | B | A | B | C |
| 人员 6 | 283 | (3) | (4) | B | D | A | B |
| 人员 7 | 280 | (4) | (1) | A | B | C | B |
| 人员 8 | 280 | (2) | (4) | B | A | A | C |
| 人员 9 | 280 | (1) | (3) | B | B | A | B |
| 人员 10 | 280 | (3) | (1) | D | B | A | C |
| 人员 11 | 278 | (4) | (1) | D | C | B | A |
| 人员 12 | 277 | (3) | (4) | A | B | C | A |
| 人员 13 | 275 | (2) | (1) | B | C | D | A |
| 人员 14 | 275 | (1) | (3) | D | B | A | B |
| 人员 15 | 274 | (1) | (4) | A | B | C | B |
| 人员 16 | 273 | (4) | (1) | B | A | B | C |

表 11-4 用人部门的基本情况以及对公务员的期望要求

| 用人部门 | 工作类别 | 各用人部门的基本情况 | | | | | 各部门对公务员特长的期望要求 | | | |
|------|------|------------|------|------|------|------|----------------|--------|--------|------|
| | | 福利待遇 | 工作条件 | 劳动强度 | 晋升机会 | 深造机会 | 专业知识面 | 认识理解能力 | 灵活应变能力 | 表达能力 |
| 部门 1 | (1) | 优 | 优 | 中 | 多 | 少 | B | A | C | A |
| 部门 2 | (2) | 中 | 优 | 大 | 多 | 少 | A | B | B | C |
| 部门 3 | (2) | 中 | 优 | 中 | 少 | 多 | | | | |
| 部门 4 | (3) | 优 | 差 | 大 | 多 | 多 | C | C | A | A |
| 部门 5 | (3) | 优 | 中 | 中 | 中 | 中 | | | | |
| 部门 6 | (4) | 中 | 中 | 中 | 中 | 多 | C | B | B | A |
| 部门 7 | (4) | 优 | 中 | 大 | 少 | 多 | | | | |

11.6.2 问题的背景与分析

目前,随着我国改革开放的不断深入和《国家公务员暂行条例》的颁布实施,几乎所有的国家机关和各省、市政府机关,以及公共事业单位等都公开面向社会招聘公务员,或工作人员.尤其是面向大中专院校的毕业生招聘活动非常普遍.一般都是采取“初试+复试+面试”的择优录用方法,特别是根据用人单位的工作性质,复试和面试在招聘录取工作中占有突出的地位.同时注意到,为了提高公务员队员素质和水平,虽然学历是反映一个人的素质和水平的一个方面,但也不能完全反映一个人的综合能力.对每个人来说,一般都各有所长,为此,如何针对应聘人员的基本素质、个人的特长和兴趣爱好,择优录用一些综合素质好、综合能力强、热爱本职工作、有专业特长的专门人才充实公务员队伍,把好人才的入口关,这在现实工作中是非常值得研究的问题.

在招聘公务员的复试过程中,如何综合专家组的意见、应聘者的不同条件和用人部门的需求做出合理的录用分配方案,这是首先需要解决的问题.当然,“多数原则”是常用的一种方法,但是,在这个问题上“多数原则”未必一定是“最好”的,因为这里有一个共性和个性的关系问题,不同的人有不同的看法和选择,怎么选择,如何兼顾考虑各方面的意见是值得研究的问题.

对于问题(1):在不考虑应聘人员的个人意愿的情况下,择优按需录用 8 名公务员.“择优”就是综合考虑所有应聘者的初试和复试的成绩来选优;“按需”就是根据用人部门的需求,即各用人部门对应聘人员的要求和评价来选择录用.而这里复试成绩没有明确给定具体分数,仅仅是专家组给出的主观评价分,为此,首先应根据专家组的评价给出一个复试分数,然后,综合考虑初试、复试分数

和用人部门的评价来确定录取名单,并按需分配给各用人部门.

对于问题(2):在充分考虑应聘人员的个人意愿的情况下,择优录用8名公务员,并按需求分配给7个用人部门.公务员和用人部门的基本情况都是透明的,在双方都是相互了解的前提下为双方做出选择方案.事实上,每一个部门对所需人才都一个期望要求,即可以认为每一个部门对每一个要聘用的公务员都有一个实际的“满意度”;同样的,每一个公务员根据自己意愿对各部门也都有一个期望“满意度”,由此根据双方的“满意度”,来选取使双方“满意度”最大的录用分配方案.

对于问题(3)是问题(1)和问题(2)的方法直接推广到一般情况就可以了.

11.6.3 模型的假设与符号说明

1. 模型的假设

- (1) 专家组对应聘者的评价是公正的;
- (2) 题中所给各部门和应聘者的相关数据都是透明的,即双方都是知道的;
- (3) 应聘者的4项特长指标在综合评价中的地位是等同的;
- (4) 用人部门的五项基本条件对公务员的影响地位是同等的.

2. 符号说明

A_j 表示第 j 个应聘者的初试得分; B_j 表示第 j 个应聘者的复试得分; C_j 表示第 j 个应聘者的最后综合得分; S_{ij} 表示第 i 个部门对第 j 个应聘者的综合满意度; T_{ji} 表示第 j 个应聘者对第 i 个部门的综合满意度; ST_{ji} 表示第 j 个应聘者与第 i 个部门的相互综合满意度; 其中 $i=1,2,\dots,7; j=1,2,\dots,16$.

11.6.4 模型的准备

1. 应聘者复试成绩的量化

首先,对专家组所给出的每一个应聘者4项条件的评分进行量化处理,从而给出每个应聘者的复试得分.注意到,专家组对应聘者的4项条件评分为A,B,C,D四个等级,不妨设相应的评语集为{很好,好,一般,差},对应的数值为5,4,3,2.根据实际情况取偏大型柯西分布隶属函数

$$f(x) = \begin{cases} [1 + \alpha(x - \beta)^{-2}]^{-1}, & 1 \leq x \leq 3, \\ a \ln x + b, & 3 < x \leq 5, \end{cases} \quad (11.6)$$

其中 α, β, a, b 为待定常数.实际上,当评价为“很好”时,则隶属度为1,即 $f(5) = 1$; 当评价为“一般”时,则隶属度为0.8,即 $f(3) = 0.8$; 当评价为“很差”时(在这里没有此评价),则认为隶属度为0.01,即 $f(1) = 0.01$. 于是,可以确定出 $\alpha = 1.1086, \beta = 0.8942, a = 0.3915, b = 0.3699$. 将其代入(11.6)式可得隶属函数为

$$f(x) = \begin{cases} [1 + 1.1086(x - 0.8942)^{-2}]^{-1}, & 1 \leq x \leq 3, \\ 0.3915 \ln x + 0.3699, & 3 < x \leq 5, \end{cases}$$

其图形如图 11-1.

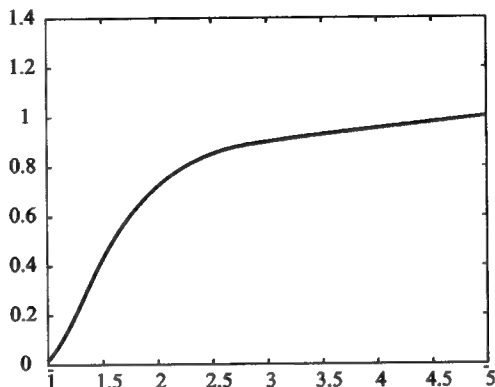


图 11-1 隶属函数的图形

经计算得 $f(2) = 0.5245$, $f(4) = 0.9126$, 则专家组对应聘者各单项指标的评价 $\{A, B, C, D\} = \{\text{很好, 好, 一般, 差}\}$ 的量化值为 $(1, 0.9126, 0.8, 0.5245)$. 根据表 11-3 的数据可以得到专家组对每一个应聘者的 4 项条件的评价指标值. 例如: 专家组对第 1 个应聘者的评价为 (A, A, B, B) , 则其指标量化值为 $(1, 1, 0.9126, 0.9126)$. 专家组对于 16 个应聘者都有相应的评价量化值, 即得到一个评价矩阵, 记为 $R = (r_{ji})_{16 \times 4}$. 由假设(3), 应聘者的 4 项条件在综合评价中的地位是同等的, 则 16 个应聘者的综合复试得分可以表示为

$$B_j = \frac{1}{4} \sum_{i=1}^4 r_{ji} \quad (j = 1, 2, \dots, 16). \quad (11.7)$$

经计算, 16 名应聘者的复试分数如下表 11-5.

表 11-5 应聘者的综合复试成绩

| 应聘者 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|------|---------|---------|---------|---------|---------|---------|---------|---------|
| 复试分数 | 0.956 3 | 0.928 2 | 0.809 3 | 0.934 5 | 0.906 3 | 0.837 4 | 0.906 3 | 0.928 2 |
| 应聘者 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 复试分数 | 0.934 5 | 0.809 3 | 0.809 3 | 0.928 2 | 0.809 3 | 0.837 4 | 0.906 3 | 0.906 3 |

2. 初试分数与复试分数的规范化

为了便于将初试分数与复试分数做统一的比较, 首先分别用极差规范化方法作相应的规范化处理. 初试得分的规范化:

$$A'_j = \frac{A_j - \min_{1 \leq j \leq 16} A_j}{\max_{1 \leq j \leq 16} A_j - \min_{1 \leq j \leq 16} A_j} = \frac{A_j - 273}{290 - 273} \quad (j = 1, 2, \dots, 16).$$

复试得分的规范化:

$$B'_j = \frac{B_j - \min_{1 \leq j \leq 16} B_j}{\max_{1 \leq j \leq 16} B_j - \min_{1 \leq j \leq 16} B_j} = \frac{B_j - 0.8093}{0.9653 - 0.8093} \quad (j = 1, 2, \dots, 16).$$

经计算可以得到具体的结果.

3. 确定应聘人员的综合分数

不同的用人单位对待初试和复试成绩的重视程度可能会不同,在这里用参数 α ($0 < \alpha \leq 1$) 表示用人单位对初试成绩的重视程度的差异,即取初试分数和复试分数的加权和作为应聘者的综合分数,则第 j 个应聘者的综合分数为

$$C_j = \alpha A'_j + (1 - \alpha) B'_j \quad (0 \leq \alpha < 1; j = 1, 2, \dots, 16). \quad (11.8)$$

由实际数据,取适当的参数 α ($0 \leq \alpha < 1$) 可以计算出每一个应聘者的最后综合得分,根据实际需要可以分别对 $\alpha = 0.4, 0.5, 0.6, 0.7$ 来计算. 在这里不妨取 $\alpha = 0.5$, 则可以得到 16 名应聘人员的综合得分及排序如表 11-6 所示.

表 11-6 应聘者的综合得分及排序

| 应聘者 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|------|---------|---------|---------|---------|---------|---------|---------|---------|
| 综合分数 | 1 | 0.845 4 | 0.441 2 | 0.778 7 | 0.624 1 | 0.389 9 | 0.535 8 | 0.610 1 |
| 排序 | 1 | 2 | 9 | 3 | 5 | 10 | 7 | 6 |
| 应聘者 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 综合分数 | 0.631 6 | 0.205 9 | 0.147 1 | 0.521 9 | 0.058 8 | 0.154 6 | 0.359 4 | 0.330 0 |
| 排序 | 4 | 13 | 15 | 8 | 16 | 14 | 11 | 12 |

11.6.5 模型的建立与求解

问题(1):

首先注意到,作为用人单位一般不会太看重应聘人员之间初试分数的少量差异,可能更注重应聘者的特长,因此,用人单位评价一个应聘者主要依据四个方面特长. 根据每个用人部门的期望要求条件和每个应聘者的实际条件(专家组的评价)的差异,则每个用人部门客观地对各个应聘者都存在一个相应的评价指标,或称为“满意度”.

从心理学的角度来分析,每一个用人部门对应聘者的每一项指标都有一个期望“满意度”,即反映用人部门对某项指标的要求与应聘者实际水平差异的程度. 通常认为用人部门对应聘者的某项指标的满意程度可以分为“很不满意、不满意、不太满意、基本满意、比较满意、满意、很满意”七个等级,即构成了评语集

$V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$, 并赋相应的数值 1, 2, 3, 4, 5, 6, 7.

当应聘者的某项指标等级与用人部门相应的要求一致时, 则认为用人部门为基本满意, 即满意程度为 v_4 ; 当应聘者的某项指标等级比用人部门相应的要求高一级时, 则用人部门的满意度上升一级, 即满意程度为 v_5 ; 当应聘者的某项指标等级与用人部门相应的要求低一级时, 则用人部门的满意度下降一级, 即满意程度为 v_3 ; 依次类推, 则可以得到用人部门对应聘者的满意程度的关系如图 11-2. 其中列表示应聘者的指标等级,

行表示用人部门的要求等级. 由此可以计算出每一个用人部门对每一个应聘者各项指标的满意程度. 例如: 专家组对应聘者 1 的评价指标集为 $\{A, A, B, B\}$, 部门 1 的期望要求指标集为 $\{B, A, C, A\}$, 则部门 1 对应聘者 1 的满意程度为 (v_5, v_4, v_5, v_3) .

| | A | B | C | D |
|---|-------|-------|-------|-------|
| A | v_4 | v_3 | v_2 | v_1 |
| B | v_5 | v_4 | v_3 | v_2 |
| C | v_6 | v_5 | v_4 | v_3 |
| D | v_7 | v_6 | v_5 | v_4 |

图 11-2 满意程度的关系图

为了得到“满意度”的量化指标, 注意到, 人们对不满意程度的敏感远远大于对满意程度的敏感, 即用人部门对应聘者的满意程度降低一级可能导致用人部门极大的抱怨, 但对满意程度增加一级只能引起满意程度的少量增长. 根据这样一个基本事实, 则可以取近似的偏大型柯西分布隶属函数

$$f(x) = \begin{cases} [1 + \alpha(x - \beta)^{-2}]^{-1}, & 1 \leq x \leq 4, \\ a \ln x + b, & 4 < x \leq 7, \end{cases}$$

其中 α, β, a, b 为待定常数. 实际上, 当“很满意”时, 则“满意度”的量化值为 1, 即 $f(7) = 1$; 当“基本满意”时, 则“满意度”的量化值为 0.8, 即 $f(4) = 0.8$; 当“很不满意”时, 则“满意度”的量化值为 0.01, 即 $f(1) = 0.01$. 于是, 可以确定出 $\alpha = 2.4944, \beta = 0.8413, a = 0.1787, b = 0.6523$. 故

$$f(x) = \begin{cases} [1 + 2.4944(x - 0.8413)^{-2}]^{-1}, & 1 \leq x \leq 4, \\ 0.1787 \ln x + 0.6523, & 4 < x \leq 7. \end{cases}$$

经计算得 $f(2) = 0.3499, f(3) = 0.6514, f(5) = 0.9399, f(6) = 0.9725$, 则用人部门对应聘者各单项指标的评语集 $\{v_1, v_2, v_3, v_4, v_5, v_6, v_7\}$ 的量化值为 $(0.01, 0.3499, 0.6514, 0.8, 0.9399, 0.9725, 1)$. 根据专家组对 16 名应聘者四项特长评分 (表 11-3) 和 7 个部门的期望要求 (表 11-4), 则可以分别计算得到每一个部门对每一个应聘者的各单项指标的满意度的量化值, 分别记为

$$(S_{ij}^{(1)}, S_{ij}^{(2)}, S_{ij}^{(3)}, S_{ij}^{(4)}) \quad (i = 1, 2, \dots, 7; j = 1, 2, \dots, 16).$$

例如, 用人部门 1 对应聘人员 1 的单项指标的满意程度为 (v_5, v_4, v_5, v_3) , 其量化值为

$(S_{11}^{(1)}, S_{11}^{(2)}, S_{11}^{(3)}, S_{11}^{(4)}) = (0.939\ 9, 0.8, 0.939\ 9, 0.651\ 4).$

由假设(3),应聘者的4项特长指标在用人单位对应聘者的综合评价中有同等的地位,为此可取第*i*个部门对第*j*个应聘者的综合评分为

$$S_{ij} = \frac{1}{4} \sum_{l=1}^4 S_{ij}^{(l)} \quad (i = 1, 2, \cdots, 7; j = 1, 2, \cdots, 16).$$
 (11.9)

具体计算结果及排序如表11-7.

表 11-7 各用人单位对应聘者的综合评分及排序

| | | | | | | | | |
|-------------------|---------|---------|---------|---------|---------|---------|---------|---------|
| 应聘者 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 部门 1 的评分 与排序 | 0.832 8 | 0.728 4 | 0.650 3 | 0.795 6 | 0.722 5 | 0.608 5 | 0.760 7 | 0.730 6 |
| | 1 | 8 | 12 | 3 | 9 | 15 | 5 | 7 |
| 部门 2,3 的评分 与排序 | 0.870 0 | 0.835 0 | 0.685 3 | 0.835 0 | 0.797 8 | 0.720 3 | 0.797 8 | 0.832 8 |
| | 1 | 2 | 12 | 3 | 7 | 11 | 8 | 4 |
| 部门 4,5 的评分 与排序 | 0.812 0 | 0.765 6 | 0.568 1 | 0.803 8 | 0.728 4 | 0.760 7 | 0.728 4 | 0.765 6 |
| | 2 | 4 | 16 | 3 | 9 | 7 | 10 | 5 |
| 部门 6,7 的评分 与排序 | 0.841 0 | 0.730 6 | 0.644 9 | 0.806 0 | 0.757 4 | 0.720 3 | 0.768 8 | 0.792 4 |
| | 1 | 11 | 16 | 3 | 9 | 13 | 6 | 5 |
| 应聘者 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| 部门 1 的评分 与排序 | 0.768 8 | 0.580 9 | 0.609 9 | 0.797 8 | 0.650 3 | 0.656 3 | 0.760 7 | 0.722 5 |
| | 4 | 16 | 14 | 2 | 13 | 11 | 6 | 10 |
| 部门 2,3 的评分 与排序 | 0.832 8 | 0.637 5 | 0.608 5 | 0.806 0 | 0.656 3 | 0.672 4 | 0.797 8 | 0.797 8 |
| | 5 | 15 | 16 | 6 | 14 | 13 | 9 | 10 |
| 部门 4,5 的评分 与排序 | 0.832 8 | 0.685 3 | 0.725 7 | 0.765 6 | 0.637 5 | 0.760 7 | 0.728 4 | 0.728 4 |
| | 1 | 14 | 13 | 6 | 15 | 8 | 11 | 12 |
| 部门 6,7 的评分 与排序 | 0.832 8 | 0.685 3 | 0.725 7 | 0.806 0 | 0.685 3 | 0.760 7 | 0.768 8 | 0.757 4 |
| | 2 | 14 | 12 | 4 | 15 | 8 | 7 | 10 |

根据“择优按需录用”的原则,来确定录用分配方案.“择优”就是选择综合分数较高者,“按需”就是录取分配方案使得用人单位的评分尽量高.为此,我们采用下面的录取及分配方案:

用 x_{ij} 表示决策变量,即

$$x_{ij} = \begin{cases} 1, & \text{当录用第 } j \text{ 个应聘者,并将其分配给第 } i \text{ 个部门时,} \\ 0, & \text{当不录用第 } j \text{ 个应聘者,或不分配给第 } i \text{ 个部门时} \end{cases} \quad \begin{matrix} (i = 1, 2, \cdots, 7; \\ j = 1, 2, \cdots, 16) \end{matrix}.$$

于是问题就转化为求下面的优化问题的解:

$$\begin{aligned} \max z = & \sum_{i=1}^7 \left(\sum_{j=1}^{16} C_j x_{ij} + \sum_{j=1}^{16} S_{ij} x_{ij} \right), \\ \text{s. t. } & \begin{cases} \sum_{i=1}^7 \sum_{j=1}^{16} x_{ij} = 8, \\ \sum_{i=1}^7 x_{ij} \leq 1 \quad (j = 1, 2, \dots, 16), \\ 1 \leq \sum_{j=1}^{16} x_{ij} \leq 2 \quad (i = 1, 2, \dots, 7), \\ x_{ij} = 0 \text{ 或 } 1 \quad (i = 1, 2, \dots, 7; j = 1, 2, \dots, 16). \end{cases} \end{aligned} \quad (11.10)$$

求解可以得到录用分配方案,根据匈牙利算法用 Matlab 编程求得结果如表 11-8.

表 11-8 录用及分配方案

| 部门 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------|---------|------------------|---------|---------|---------|---------|---------|
| 应聘者 | 1 | 2, 5 | 8 | 9 | 4 | 7 | 12 |
| 综合分数 | 1 | 0.845 4, 0.624 1 | 0.610 1 | 0.631 6 | 0.778 7 | 0.535 9 | 0.521 9 |
| 部门评分 | 0.832 8 | 0.835 0, 0.797 8 | 0.832 8 | 0.832 8 | 0.803 8 | 0.768 8 | 0.806 0 |

问题(2):

在充分考虑应聘人员的意愿和用人部门的期望要求的情况下,寻求更好的录用分配方案.应聘人员的意愿有两个方面:对用人部门的工作类别的选择意愿和对用人部门的基本情况的看法,即可用应聘人员对用人部门的综合满意度来表示;用人部门对应聘人员的期望要求也用满意度来表示.一个好的录用分配方案应该是使得二者的满意度都尽量的高.

(1) 确定用人部门对应聘者的满意度

用人部门对所有应聘人员的满意度与问题(1)中的(11.9)式相同,即第 i 个部门对第 j 个应聘人员的 4 项条件的综合评价满意度为

$$S_{ij} = \frac{1}{4} \sum_{t=1}^4 S_{ij}^{(t)} \quad (i = 1, 2, \dots, 7; j = 1, 2, \dots, 16).$$

(2) 确定应聘者对用人部门的满意度

应聘者对用人部门的满意度主要与用人部门的基本情况有关,同时考虑到应聘者所喜好的工作类别,在评价用人部门时一定会偏向于自己的喜好,即工作类别也是决定应聘者选择部门的一个因素.因此,影响应聘者对用人部门的满意度的有五项指标:福利待遇、工作条件、劳动强度、晋升机会和深造机会.

对工作类别来说,主要看是否符合自己想从事的工作,符合第一、二志愿的

分别为“满意、基本满意”,不符合志愿的为“不满意”,即{满意,基本满意,不满意}. 实际中根据人们对待工作类别志愿的敏感程度的心理变化,在这里取隶属函数为 $f(x) = b \ln(a - x)$,并要求 $f(1) = 1, f(3) = 0$,即符合第一志愿时,满意度为1,不符合任一个志愿时满意度为0,简单计算解得 $a = 4, b = 0.9102$,即 $f(x) = 0.9102 \ln(4 - x)$. 于是当用人部门的工作类别符合应聘者的第二志愿时的满意度分别为 $f(2) = 0.6309$,即得到评语集{满意,基本满意,不满意}的量化值为 $(1, 0.6309, 0)$. 这样每一个应聘者对每一个用人部门都有一个满意度权值 w_{ji} ($i = 1, 2, \dots, 7; j = 1, 2, \dots, 16$),即满足第一志愿取权为1,满足第二志愿取权值为0.6309,不满足志愿取权值为0.

对于反映用人部门基本情况五项指标都可分为“优中差,或小中大、多中少”三个等级,应聘者对各部门的评语集也为三个等级,即{满意,基本满意,不满意},类似于上面确定用人部门对应聘者的满意度的方法.

首先确定用人部门基本情况的客观指标值:应聘者对7个部门的五项指标中的“优、小、多”级别认为很满意,其隶属度为1;“中”级别认为满意,其隶属度为0.6;“差、大、少”级别认为不满意,其隶属度为0.1. 由表11-4的实际数据可得应聘者对每个部门的各单项指标的满意度量化值,即用人部门的客观水平的评价值 $T_i = (T_{i1}, T_{i2}, T_{i3}, T_{i4}, T_{i5})$ ($i = 1, 2, \dots, 7$),具体结果如表11-9.

表 11-9 用人部门的基本情况的量化指标

| 部门 指标 | 部门 1 T_{1k} | 部门 2 T_{2k} | 部门 3 T_{3k} | 部门 4 T_{4k} | 部门 5 T_{5k} | 部门 6 T_{6k} | 部门 7 T_{7k} |
|----------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| 1 | 1 | 0.6 | 0.6 | 1 | 1 | 0.6 | 1 |
| 2 | 1 | 1 | 1 | 0.1 | 0.6 | 0.6 | 0.6 |
| 3 | 0.6 | 0.1 | 0.6 | 0.1 | 0.6 | 0.6 | 0.1 |
| 4 | 1 | 1 | 0.1 | 1 | 0.6 | 0.6 | 0.1 |
| 5 | 0.1 | 0.1 | 1 | 1 | 0.6 | 1 | 1 |

于是,每一个应聘者对每一个部门的五个单项指标的满意度应为该部门的客观水平评价与应聘者对该部门的满意度权值 w_{ji} ($i = 1, 2, \dots, 7; j = 1, 2, \dots, 16$)的乘积,即

$$\begin{aligned} \bar{T}_{ji} &= w_{ji} \cdot (T_{i1}, T_{i2}, T_{i3}, T_{i4}, T_{i5}) = (T_{ji}^{(1)}, T_{ji}^{(2)}, T_{ji}^{(3)}, T_{ji}^{(4)}, T_{ji}^{(5)}) \\ &\quad (i = 1, 2, \dots, 7; j = 1, 2, \dots, 16). \end{aligned}$$

例如,应聘者1对部门5的单项指标的满意度为

$$\begin{aligned} \bar{T}_{15} &= (T_{15}^{(1)}, T_{15}^{(2)}, T_{15}^{(3)}, T_{15}^{(4)}, T_{15}^{(5)}) = 0.6309 \cdot (1, 0.6, 0.6, 0.6, 0.6) \\ &= (0.6309, 0.3785, 0.3785, 0.3785, 0.3785). \end{aligned}$$

由假设(3),用人部门的五项指标在应聘者对用人部门的综合评价中有同等的地位,为此可取第 j 个应聘者对第 i 个部门的综合评价满意度为

$$T_{ji} = \frac{1}{5} \sum_{k=1}^5 T_{ji}^{(k)} \quad (i = 1, 2, \dots, 7; j = 1, 2, \dots, 16). \quad (11.11)$$

(3) 确定双方的相互综合满意度

根据上面的讨论,每一个用人部门与每一个应聘者之间都有相应的单方面的满意度,双方的相互满意度应由各自的满意度来确定,在此,取双方各自满意度的几何平均值为双方相互综合满意度,即

$$ST_{ij} = \sqrt{S_{ij} \cdot T_{ji}} \quad (i = 1, 2, \dots, 7; j = 1, 2, \dots, 16). \quad (11.12)$$

(4) 确定合理的录用分配方案

最优的录用分配方案应该是使得所有用人部门和录用的公务员之间的相互综合满意度之和最大. 设决策变量

$$x_{ij} = \begin{cases} 1, & \text{当录用第 } j \text{ 个应聘者,并分配到第 } i \text{ 个部门时,} \\ 0, & \text{当不录用第 } j \text{ 个应聘者,或不分配到第 } i \text{ 个部门时} \end{cases} \quad \begin{cases} i = 1, 2, \dots, 7; \\ j = 1, 2, \dots, 16 \end{cases}$$

于是问题可以归结为下面的线性 0-1 规划问题

$$\begin{aligned} \max z = & \sum_{i=1}^7 \sum_{j=1}^{16} ST_{ij} \cdot x_{ij}, \\ \text{s. t. } & \begin{cases} \sum_{i=1}^7 \sum_{j=1}^{16} x_{ij} = 8, \\ \sum_{i=1}^7 x_{ij} = 1 \quad (j = 1, 2, \dots, 16), \\ 1 \leq \sum_{j=1}^{16} x_{ij} \leq 2 \quad (j = 1, 2, \dots, 7), \\ x_{11} = x_{14} = x_{15} = x_{16} = x_{18} = x_{1,12} = 0, \\ x_{22} = x_{24} = x_{26} = x_{27} = x_{29} = x_{2,10} = x_{2,11} = x_{2,12} = x_{2,14} \\ \quad = x_{2,15} = x_{2,16} = 0, i = 2, 3, \\ x_{33} = x_{37} = x_{38} = x_{3,11} = x_{3,13} = x_{3,15} = x_{3,16} = 0, i = 4, 5, \\ x_{41} = x_{42} = x_{43} = x_{45} = x_{49} = x_{4,10} = x_{4,13} = x_{4,14} = 0, i = 6, 7, \\ x_{ij} = 0 \text{ 或 } 1 \quad (i = 1, 2, \dots, 7; j = 1, 2, \dots, 16), \end{cases} \end{aligned} \quad (11.13)$$

其中第 1 个条件是当且仅当录取 8 名,第 2 个条件是限制一个应聘者仅允许分配一个部门,第 3 个条件是保证每一个用人部门至少录用 1 名、至多录用 2 名公务员,第 4—7 个条件是应聘者不可能分配的部门约束.

该模型为一个线性 0-1 规划,用 MATLAB 编程求得录用分配方案如表

11-10, 总满意度 $z = 5.7631$.

表 11-10 最终的录用分配方案

| 部门序号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|-------|------------------|---------|---------|---------|---------|---------|---------|
| 应聘者序号 | 9, 15 | 8 | 1 | 12 | 2 | 4 | 7 |
| 综合满意度 | 0.754 3, 0.750 3 | 0.682 9 | 0.757 7 | 0.700 0 | 0.721 5 | 0.740 3 | 0.656 1 |

问题(3):

对于 N 个应聘人员和 $M (M < N)$ 个用人单位的情况, 如上的方法都是实用的, 只是两个优化模型(11.10)和(11.13)的规模将会增大, 给求解带来一定的困难. 实际中用人单位的个数 M 不会太大, 当应聘人员的个数 N 大到一定的程度时, 可以分步处理: 先根据应聘人的综合分数和用人部门的评价分数择优确定录用名单, 然后再“按需”分配.

对问题(1)而言, 取所有应聘人员综合分数与用人部门综合评分的均值, 即由(11.8)式和(11.9)式得

$$\bar{C} = \frac{1}{N} \sum_{j=1}^N C_j \text{ 和 } \bar{S} = \frac{1}{NM} \sum_{i=1}^M \sum_{j=1}^N S_{ij}.$$

对于满足 $C_j < \bar{C}$ 或 $\frac{1}{M} \sum_{i=1}^M S_{ij} < \bar{S} (j = 1, 2, \dots, N)$ 的应聘人员淘汰掉, 对剩下的应聘者重新编号, 再用上述的方法求解, 确定录用分配方案. 如果剩下的人数仍然很多, 则可以做类似地进一步择优.

对于问题(2)处理的方法类似, 只是根据应聘人员的综合分数(11.8)和双方综合满意度(11.12)来选优.

11.7 参考案例与参考文献

1. 参考案例

- (1) 学生选课问题——文献[1]:109—112
- (2) 会议分组问题——文献[2]:170—185
- (3) 铁路平板车的装货问题——文献[3]:292—294
- (4) 飞机排队问题——文献[3]:297—298
- (5) 截断切割问题——文献[4]:286—301
- (6) 公司的战略重组问题——文献[5]:198—203

2. 参考文献

- [1] 姜启源. 数学模型. 第三版. 北京: 高等教育出版社, 2003
- [2] 叶其孝. 大学生数学建模竞赛辅导教材(四). 长沙: 湖南教育出版

社,1997

[3] 李尚志. 数学建模数学建模竞赛教程. 南京:江苏教育出版社,1996

[4] 汪国强. 数学建模优秀案例选编. 广州:华南理工大学出版社,1998

[5] 全国大学生数学建模竞赛组委会. 全国大学生数学建模竞赛优秀论文汇编. 北京:中国物价出版社,2002

[6] 杨超等. 运筹学. 北京:科学出版社,2004

[7] 《运筹学》教材编写组. 运筹学. 修订版. 北京:清华大学出版社,1990

[8] 韩中庚. 招聘公务员问题的优化模型与评述. 工程数学学报. 2004, 21(7):147—154

第十二章 非线性规划方法

实际中许多较复杂的问题都可归结为一个非线性规划问题,即如果目标函数和约束条件中包含有非线性函数,则这样的规划问题称为非线性规划问题. 解决这类问题要用非线性的方法,但一般说来,解决非线性的问题要比解决线性的问题困难得多,不像线性规划有适用于一般情况的单纯形法. 我们知道线性规划的可行域一般是一个凸集,如果线性规划存在最优解的,则其最优解一定在可行域的边界上达到(特别是在可行域的顶点上达到). 而对于非线性规划,即便存在最优解,则可以在其可行域的任何点达到. 因此,对于非线性规划问题到目前为止还没有一种适用于一般情况的求解方法,现有各种方法都有各自特定的适用范围,为此,这也是一个正处在研究发展中的学科领域.

12.1 非线性规划的基本概念

12.1.1 非线性规划问题的数学模型

非线性规划的一般模型为

$$\begin{cases} \min f(x_1, x_2, \dots, x_n), \\ h_i(x_1, x_2, \dots, x_n) = 0, i = 1, 2, \dots, m, \\ g_j(x_1, x_2, \dots, x_n) \geq 0, j = 1, 2, \dots, l. \end{cases} \quad (12.1)$$

若记 $X = (x_1, x_2, \dots, x_n)^T \in E^n$ 是 n 维欧氏空间中的向量(点),则其模型为

$$\begin{cases} \min f(X), \\ h_i(X) = 0, i = 1, 2, \dots, m, \\ g_j(X) \geq 0, j = 1, 2, \dots, l. \end{cases} \quad (12.2)$$

说明:

(1) 若目标函数为最大化问题,由 $\max f(X) = -\min[-f(X)]$,令 $F(X) = -f(X)$,则 $\min F(X) = -\max f(X)$;

(2) 若约束条件为 $g_j(X) \leq 0$,则 $-g_j(X) \geq 0$;

(3) $h_i(X) = 0 \Leftrightarrow h_i(X) \geq 0$ 且 $-h_i(X) \geq 0$.

于是可将非线性规划的一般模型写成如下形式:

$$\begin{cases} \min f(X), \\ g_j(X) \geq 0, j = 1, 2, \dots, m. \end{cases} \quad (12.3)$$

12.1.2 几种特殊情况

1. 无约束的非线性规划

当问题无约束条件时,则此问题称为无约束的非线性规划问题,即求多元函数的极值问题.一般模型为

$$\begin{cases} \min_{X \in R} f(X), \\ X \geq 0. \end{cases} \quad (12.4)$$

2. 二次规划

如果目标函数是 X 的二次函数,约束条件都是线性的,则称此规划为二次规划.二次规划的一般模型为

$$\begin{cases} \min f(X) = \sum_{j=1}^n c_j x_j + \sum_{j=1}^n \sum_{k=1}^n c_{jk} x_j x_k, \\ \sum_{j=1}^n a_{ij} x_j + b_i \geq 0, i = 1, 2, \dots, m, \\ x_j \geq 0, c_{jk} = c_{kj}, j, k = 1, 2, \dots, n. \end{cases} \quad (12.5)$$

3. 凸规划

当模型(12.3)中的目标函数 $f(X)$ 为凸函数, $g_j(X)$ ($j = 1, 2, \dots, m$) 均为凹函数(即 $-g_j(X)$ 为凸函数),则这样的非线性规划称为凸规划.

12.1.3 多元函数的极值

1. 局部极值与全局极值

设 $f(X)$ 在区域 $R \subset E^n$ 上有定义,对于 $X^* \in R$,如果存在某个 $\varepsilon > 0$,使对满足 $\|X - X^*\| < \varepsilon$ 的所有 $X \in R$ 都有 $f(X) \geq f(X^*)$ (或 $f(X) \leq f(X^*)$),则称 X^* 为 $f(X)$ 在 R 上局部极小(或大)点, $f(X^*)$ 为局部极小(或大)值.

如果对所有满足 $\|X - X^*\| < \varepsilon$ 的 $X \in R$ ($X \neq X^*$) 都有 $f(X) > f(X^*)$ (或 $f(X) < f(X^*)$),则称 X^* 为 $f(X)$ 在 R 上严格局部极小(或大)点, $f(X^*)$ 为严格局部极小(或大)值.

如果对所有 $X \in R$ 都有 $f(X) \geq f(X^*)$ (或 $f(X) \leq f(X^*)$),则称 X^* 为 $f(X)$ 在 R 上全局极小(或大)点, $f(X^*)$ 为全局极小(或大)值.

如果对所有 $X \in R$ ($X \neq X^*$) 都有 $f(X) > f(X^*)$ (或 $f(X) < f(X^*)$),则称 X^* 为 $f(X)$ 在 R 上严格全局极小(或大)点, $f(X^*)$ 为严格全局极小(或大)值.

2. 极值存在的条件

定理 12.1 (必要条件) 设 $f(X)$ 在开集 $R \subset E^n$ 内有一阶连续的偏导数, 且在点 $X^* \in R$ 取得局部极值, 则梯度向量 $\nabla f(X^*) = 0$, 其中

$$\nabla f(X^*) = \left(\frac{\partial f(X^*)}{\partial x_1}, \frac{\partial f(X^*)}{\partial x_2}, \dots, \frac{\partial f(X^*)}{\partial x_n} \right)^T.$$

而梯度 $\nabla f(X)$ 的方向是函数 $f(X)$ 的等值面 (或等值线) 的法线方向, 即是函数值增加最快的方向, 而负方向是函数值减少最快的方向.

定理 12.2 (充分条件) 设 $f(X)$ 在开集 $R \subset E^n$ 内有二阶连续的偏导数, 若有 $X^* \in R$ 使 $\nabla f(X^*) = 0$, 且对任何非零向量 $X \in R$ 有 $X^T H(X^*) X > 0$, 则 X^* 为 $f(X)$ 的严格局部极小点. 其中 $H(X^*)$ 为 $f(X)$ 在 X^* 点的海赛 (Hesse) 矩阵, 即

$$H(X^*) = \begin{bmatrix} \frac{\partial^2 f(X^*)}{\partial x_1^2} & \frac{\partial^2 f(X^*)}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f(X^*)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(X^*)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(X^*)}{\partial x_2^2} & \dots & \frac{\partial^2 f(X^*)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(X^*)}{\partial x_n \partial x_1} & \frac{\partial^2 f(X^*)}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f(X^*)}{\partial x_n^2} \end{bmatrix}$$

12.2 无约束非线性规划的解法

12.2.1 一般迭代法

迭代法是求解非线性规划问题的最常用的一种数值方法, 其基本思想是: 对于问题 (12.4) 而言, 给出 $f(X)$ 的极小点的初始值 $X^{(0)}$, 按某种规律计算出一系列的 $X^{(k)}$ ($k=1, 2, \dots$), 希望点列 $\{X^{(k)}\}$ 的极限 X^* 就是 $f(X)$ 的一个极小点.

现在的问题是: 如何来产生这个点列? 即如何由一个解向量 $X^{(k)}$ 求出另一个新的解向量 $X^{(k+1)}$?

实际上: 向量总是由方向和长度确定, 即向量 $X^{(k+1)}$ 总可以写成

$$X^{(k+1)} = X^{(k)} + \lambda_k P^{(k)} \quad (k = 1, 2, \dots)$$

其中 $P^{(k)}$ 为一个向量, λ_k 为一个实数, 称为步长, 即 $X^{(k+1)}$ 可由 λ_k 及 $P^{(k)}$ 惟一确定.

实际中, 各种迭代法的区别就在于寻求 λ_k 和 $P^{(k)}$ 方式的不同, 特别是方向向量 $P^{(k)}$ 的确定是问题的关键, 称为搜索方向. 选择 λ_k 和 $P^{(k)}$ 的一般原则是使目标函数在这些点列上的值逐步减小, 即

$$f(X^{(0)}) \geq f(X^{(1)}) \geq \cdots \geq f(X^{(k)}) \geq \cdots$$

为此,这种算法称为下降算法.最后要检验 $\{X^{(k)}\}$ 是否收敛于最优解,即对于给定的精度 $\varepsilon > 0$,是否有 $\|\nabla f(X^{(k+1)})\| \leq \varepsilon$,决定迭代过程是否结束.

12.2.2 一维搜索法

沿着一系列的射线方向 $P^{(k)}$ 寻求极小化点列的方法称为一维搜索法,这是一类方法.

对于确定的方向 $P^{(k)}$,在射线 $X^{(k)} + \lambda P^{(k)}$ ($\lambda \geq 0$)上选取步长 λ_k 使 $f(X^{(k)} + \lambda_k P^{(k)}) < f(X^{(k)})$,则可以确定一个新的点 $X^{(k+1)} = X^{(k)} + \lambda_k P^{(k)}$,即为沿射线 $X^{(k)} + \lambda P^{(k)}$ 求函数 $f(X)$ 的最小值的问题.即等价于求一元函数 $\phi(\lambda) = f(X^{(k)} + \lambda P^{(k)})$ 在点集 $L = \{X | X = X^{(k)} + \lambda P^{(k)}, -\infty < \lambda < \infty\}$ 上的极小点 λ_k .

一维搜索法是对某一个确定方向 $P^{(k)}$ 来进行的,现在的问题是如何选择搜索方向 $P^{(k)}$ 呢?

12.2.3 梯度法(最速下降法)

选择一个使函数值下降速度最快的方向.考虑到 $f(X)$ 在点 $X^{(k)}$ 处沿着方向 P 的方向导数为 $f_P(X^{(k)}) = \nabla f(X^{(k)})^T \cdot P$,其意义是指 $f(X)$ 在点 $X^{(k)}$ 处沿方向 P 的变化率.当 $f(X)$ 连续可微,且方向导数为负时,说明函数值沿该方向下降,方向导数越小,表明下降的速度就越快.因此,可以把 $f(X)$ 在 $X^{(k)}$ 点的方向导数最小的方向(即梯度的负方向)作为搜索方向,即令 $P^{(k)} = -\nabla f(X^{(k)})$,这就是梯度法,或最速下降法.

梯度法的计算步骤:

(1) 选定初始点 $X^{(0)}$ 和给定精度要求 $\varepsilon > 0$,令 $k=0$;

(2) 若 $\|\nabla f(X^{(k)})\| < \varepsilon$,则停止计算, $X^* = X^{(k)}$,否则令 $P^{(k)} = -\nabla f(X^{(k)})$;

(3) 在 $X^{(k)}$ 处沿方向 $P^{(k)}$ 作一维搜索得 $X^{(k+1)} = X^{(k)} + \lambda_k P^{(k)}$,令 $k=k+1$,返回第二步,直到求得最优解为止.实际上,可以求得

$$\lambda_k = \frac{\nabla f(X^{(k)})^T \cdot \nabla f(X^{(k)})}{\nabla f(X^{(k)})^T \cdot H(X^{(k)}) \cdot \nabla f(X^{(k)})}.$$

12.2.4 共轭梯度法

共轭梯度法仅适用于正定二次函数的极小值问题:

$$\min f(X) = \frac{1}{2} X^T A X + B^T X + c,$$

其中 A 为 $n \times n$ 阶实对称正定阵, $X, B \in E^n, c$ 为常数.

定义 12.1 设 A 为 $n \times n$ 阶实对称正定阵, 若对 n 维向量 P_1 和 P_2 满足 $P_1^T A P_2 = 0$, 则称向量 P_1 和 P_2 关于 A 共轭(正交).

从任意初始点 $X^{(1)}$ 和向量 $P^{(1)} = -\nabla f(X^{(1)})$ 出发, 由

$$X^{(k+1)} = X^{(k)} + \lambda_k P^{(k)}, \quad \lambda_k = \min_{\lambda} f(X^{(k)} + \lambda P^{(k)}) = -\frac{(\nabla f(X^{(k)}))^T P^{(k)}}{(P^{(k)})^T A P^{(k)}}$$

和

$$P^{(k+1)} = -\nabla f(X^{(k+1)}) + \beta_k P^{(k)}, \quad \beta_k = \frac{(P^{(k)})^T \cdot A \cdot \nabla f(X^{(k+1)})}{(P^{(k)})^T A P^{(k)}}$$

$$(k = 1, 2, \dots, n-1),$$

可以得到 $(X^{(2)}, P^{(2)}), (X^{(3)}, P^{(3)}), \dots, (X^{(n)}, P^{(n)})$. 能够证明向量 $P^{(1)}, P^{(2)}, \dots, P^{(n)}$ 是线性无关的, 且关于 A 是两两共轭的. 从而可以得到 $\nabla f(X^{(n)}) = 0$, 则 $X^{(n)}$ 为 $f(X)$ 的极小点. 这就是共轭梯度法, 其计算步骤如下:

(1) 对任意初始点 $X^{(1)} \in E^n$ 和向量 $P^{(1)} = -\nabla f(X^{(1)})$, 取 $k=1$;

(2) 若 $\nabla f(X^{(k)}) = 0$, 即得到最优解, 停止计算; 否则求

$$X^{(k+1)} = X^{(k)} + \lambda_k P^{(k)}, \quad \lambda_k = \min_{\lambda} f(X^{(k)} + \lambda P^{(k)}) = -\frac{(\nabla f(X^{(k)}))^T P^{(k)}}{(P^{(k)})^T A P^{(k)}},$$

$$P^{(k+1)} = -\nabla f(X^{(k+1)}) + \beta_k P^{(k)}, \quad \beta_k = \frac{(P^{(k)})^T \cdot A \cdot \nabla f(X^{(k+1)})}{(P^{(k)})^T A P^{(k)}}$$

$$(k = 1, 2, \dots, n-1).$$

(3) 令 $k=k+1$; 返回(2).

注: 对于一般的二阶可微函数 $f(X)$, 在每一点的局部可以近似地视为二次函数

$$f(X) \approx f(X^{(k)}) + \nabla f(X^{(k)})^T (X - X^{(k)}) + \frac{1}{2} (X - X^{(k)})^T \nabla^2 f(X^{(k)}) (X - X^{(k)}).$$

类似地可以用共轭梯度法处理.

12.2.5 牛顿(Newton)法

对于问题:

$$\min f(X) = \frac{1}{2} X^T A X + B^T X + c,$$

由 $\nabla f(X) = A X + B = 0$, 则由最优性条件 $\nabla f(X) = 0$, 当 A 为正定时, A^{-1} 存在, 于是有 $X^* = -A^{-1} B$ 为最优解.

12.2.6 拟牛顿法

对于一般的二阶可微函数 $f(X)$, 在 $X^{(k)}$ 点的局部有

$$f(X) \approx f(X^{(k)}) + \nabla f(X^{(k)})^T (X - X^{(k)}) + \frac{1}{2} (X - X^{(k)})^T \nabla^2 f(X^{(k)}) (X - X^{(k)}).$$

当海赛矩阵 $\nabla^2 f(X^{(k)})$ 正定时,也可应用上面的牛顿法,这就是拟牛顿法,其计算步骤如下:

- (1) 任取 $X^{(1)} \in E^n, k=1$;
- (2) 计算 $g_k = \nabla f(X^{(k)})$, 若 $g_k = 0$, 则停止计算, 否则计算 $H(X^{(k)}) = \nabla^2 f(X^{(k)})$, 令 $X^{(k+1)} = X^{(k)} - (H(X^{(k)}))^{-1} g_k$;
- (3) 令 $k = k+1$; 返回(2).

这种方法虽然简单,但选取初始值是比较困难的,选取不好可能不收敛.另外,对于一般的目标函数很复杂,或 X 的维数的很高时,要计算二阶导和求逆阵也是很困难的,或根本不可能.为了解决这个问题对上面的方法进行修正,即修正搜索方向,避免求二阶导数和逆矩阵,其他的都与拟牛顿法相同,这就是下面的变尺度法.

12.2.7 变尺度法

变尺度法的计算步骤如下:

- (1) 任取 $X^{(0)} \in E^n$ 和 $H^{(0)}$ (一般取 $H^{(0)} = I$ 为单位阵), 计算 $P^{(0)} = -H^{(0)} \cdot \nabla f(X^{(0)}), k=0$;
- (2) 若 $\nabla f(X^{(k)}) = 0$, 则停止计算, 否则令 $X^{(k+1)} = X^{(k)} + \lambda_k P^{(k)}$, 其中 λ_k 为最佳步长, 由 $\min_{\lambda} f(X^{(k)} + \lambda P^{(k)}) = f(X^{(k)} + \lambda_k P^{(k)})$ 确定;
- (3) 计算 $\delta_{k+1} = X^{(k+1)} - X^{(k)}, \gamma_{k+1} = \nabla f(X^{(k+1)}) - \nabla f(X^{(k)})$,

$$H^{(k+1)} = H^{(k)} + \frac{\delta_{k+1} \cdot \delta_{k+1}^T}{\delta_{k+1}^T \cdot \gamma_{k+1}} - \frac{H^{(k)} \cdot \gamma_{k+1} \cdot \gamma_{k+1}^T \cdot H^{(k)}}{\gamma_{k+1}^T \cdot H^{(k)} \cdot \gamma_{k+1}},$$

$$P^{(k+1)} = -H^{(k+1)} \cdot \nabla f(X^{(k+1)});$$
- (4) 令 $k = k+1$; 返回(2).

12.3 带有约束条件的非线性规划

12.3.1 非线性规划的最优性条件

在给出非线性规划的最优性条件之前,为了说明方便,我们首先引入两个概念.

定义 12.2 设 $X^{(0)}$ 是非线性规划问题(12.3)的一个可行解,它使得某个约束条件 $g_j(X) \geq 0$ ($1 \leq j \leq l$), 具体有下面两种情况:

- (1) 如果使 $g_j(X^{(0)}) > 0$, 则称约束条件 $g_j(X) \geq 0$ ($1 \leq j \leq l$) 是 $X^{(0)}$ 点的

无效约束(或不起作用的约束);

(2) 如果使 $g_j(X^{(0)}) = 0$, 则称约束条件 $g_j(X) \geq 0$ ($1 \leq j \leq l$) 是 $X^{(0)}$ 点的有效约束(或起作用的约束)。

实际上, 如果 $g_j(X) \geq 0$ ($1 \leq j \leq l$) 是 $X^{(0)}$ 点的无效约束, 则说明 $X^{(0)}$ 位于可行域的内部, 不在边界上, 即当 $X^{(0)}$ 有微小变化时, 此约束条件没有什么影响。而对有效约束则说明 $X^{(0)}$ 位于可行域的边界上, 即当 $X^{(0)}$ 有微小变化时, 此约束条件起着限制作用。

定义 12.3 设 $X^{(0)}$ 是非线性规划问题(12.3)的一个可行解, D 是过此点的某一个方向, 如果

(1) 存在实数 $\lambda_0 > 0$, 使对任意 $\lambda \in [0, \lambda_0]$ 均有 $X^{(0)} + \lambda D \in R$, 则称此方向 D 是 $X^{(0)}$ 点一个可行方向;

(2) 存在实数 $\lambda_0 > 0$, 使对任意 $\lambda \in [0, \lambda_0]$ 均有 $f(X^{(0)} + \lambda D) < f(X^{(0)})$, 则称此方向 D 是 $X^{(0)}$ 点一个下降方向;

(3) 方向 D 既是 $X^{(0)}$ 点的可行方向, 又是下降方向, 则称它是 $X^{(0)}$ 点可行下降方向。

实际中, 如果某个 $X^{(0)}$ 不是极小点(最优解), 就继续沿着 $X^{(0)}$ 点的可行下降方向去搜索。显然, 若 $X^{(0)}$ 点存在可行下降方向, 它就不是极小点; 另一方面, 若 $X^{(0)}$ 为极小点, 则该点就不存在可行下降方向。

下面针对非线性规划问题(12.3)给出最优性条件。

定理 12.3 (Kuhn-Tucker) 如果 X^* 是问题(12.3)的极小点, 且对点 X^* 有效约束的梯度线性无关, 则必存在向量 $\Gamma^* = (\gamma_1^*, \gamma_2^*, \dots, \gamma_m^*)^T$ 使下述条件成立:

$$\begin{cases} \nabla f(X^*) - \sum_{j=1}^m \gamma_j^* \nabla g_j(X^*) = 0, \\ \gamma_j^* g_j(X^*) = 0, j = 1, 2, \dots, m, \\ \gamma_j^* \geq 0, j = 1, 2, \dots, m. \end{cases}$$

此条件称为库恩-塔克(Kuhn-Tucker)条件, 简称为 K-T 条件。满足 K-T 条件的点称 K-T 点。

类似地, 如果 X^* 是问题(12.2)的极小点, 且对点 X^* 所有有效约束的梯度 $\nabla h_i(X^*)$ ($i = 1, 2, \dots, m$) 和 $\nabla g_j(X^*)$ ($j = 1, 2, \dots, l$) 线性无关, 则必存在向量 $\Lambda^* = (\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*)^T$ 和 $\Gamma^* = (\gamma_1^*, \gamma_2^*, \dots, \gamma_l^*)^T$ 使下面的 K-T 条件成立:

$$\begin{cases} \nabla f(X^*) - \sum_{i=1}^m \lambda_i^* \nabla h_i(X^*) - \sum_{j=1}^l \gamma_j^* \nabla g_j(X^*) = 0, \\ \gamma_j^* g_j(X^*) = 0, j = 1, 2, \dots, l, \\ \gamma_j^* \geq 0, j = 1, 2, \dots, l. \end{cases}$$

将满足 K-T 条件的点也称为 K-T 点. 其中 $\lambda_1^*, \lambda_2^*, \dots, \lambda_m^*$ 和 $\gamma_1^*, \gamma_2^*, \dots, \gamma_l^*$ 称为广义 Lagrange 乘子.

库恩-塔克条件是非线性规划最重要的理论基础, 是确定某点是否为最优解(点)的必要条件, 但一般不是充分条件, 即满足这个条件的点不一定是最优解. 但对于凸规划它一定是最优解的充要条件.

12.3.2 非线性规划的可行方向法

考虑非线性规划问题(12.3), 假设 $X^{(k)}$ 是该问题的一个可行解, 但不是最优解, 为了进一步寻找最优解在它的可行下降方向中选取其一个方向 $D^{(k)}$, 并确定最佳步长 λ_k 使得

$$\begin{cases} X^{(k+1)} = X^{(k)} + \lambda_k D^{(k)} \in R, \\ f(X^{(k+1)}) < f(X^{(k)}), \end{cases} \quad k = 0, 1, 2, \dots$$

反复进行这一过程, 直到得到满足精度要求为止, 这种方法称为可行方向法.

可行方向法的主要特点是: 因为迭代过程中所采用的搜索方向总为可行方向, 所以产生的迭代点列 $\{X^{(k)}\}$ 始终在可行域 R 内, 且目标函数值不断地单调下降. 可行方向法实际上是一类方法, 最典型的是 Zoutendijk 可行方向法.

定理 12.4 设 X^* 是问题(12.3)的一个局部极小点, 函数 $f(X)$ 和 $g_j(X)$ 在 X^* 处均可微, 则在 X^* 点不存在可行下降的方向, 从而不存在向量 D 同时满足

$$\begin{cases} \nabla f(X^*)^T D < 0, \\ \nabla g_j(X^*)^T D > 0, j = 1, 2, \dots, m. \end{cases}$$

实际上, 由

$$\begin{cases} f(X^* + \lambda D) = f(X^*) + \lambda \nabla f(X^*)^T D + o(\lambda), \\ g_j(X^* + \lambda D) = g_j(X^*) + \lambda \nabla g_j(X^*)^T D + o(\lambda), \end{cases}$$

可知这个定理的结论是显然的, 否则就与 X^* 是极小点矛盾.

Zoutendijk 可行方向法: 设 $X^{(k)}$ 点的有效约束集非空, 则 $X^{(k)}$ 点的可行下降方向 $D = (d_1, d_2, \dots, d_n)^T$ 必满足

$$\begin{cases} \nabla f(X^{(k)})^T D < 0, \\ \nabla g_j(X^{(k)})^T D > 0, j \in J, \end{cases}$$

又等价于

$$\begin{cases} \nabla f(X^{(k)})^T D \leq \eta, \\ -\nabla g_j(X^{(k)})^T D \leq \eta, j \in J, \\ \eta < 0, \end{cases}$$

其中 J 是有效约束的下标集. 此问题可以转化为求下面的线性规划问题:

$$\begin{aligned} & \min \eta, \\ & \begin{cases} \nabla f(X^{(k)})^T D \leq \eta, \\ -\nabla g_j(X^{(k)})^T D \leq \eta, j \in J, \\ -1 \leq d_i \leq 1, i = 1, 2, \dots, n, \end{cases} \end{aligned}$$

其中最后一个约束是为了求问题的有限解, 即只需要确定 D 的方向, 只要确定单位向量即可.

如果求得 $\eta = 0$, 则在 $X^{(k)}$ 点不存在可行下降方向, $X^{(k)}$ 就是 K-T 点. 如果求得 $\eta < 0$, 则可以得到可行下降方向 $D^{(k)}$. 这就是 Zoutendijk 可行方向法.

实际中, 利用 Zoutendijk 可行方向法得到可行下降方向 $D^{(k)}$ 后, 用求一维极值的方法求出最佳步长 λ_k , 则再进行下一步的迭代

$$\begin{cases} X^{(k+1)} = X^{(k)} + \lambda_k D^{(k)} \in R, \\ f(X^{(k+1)}) < f(X^{(k)}), \end{cases} k = 0, 1, 2, \dots.$$

12.4 非线性规划的制约函数法

制约函数的基本思想是: 将求解非线性规划的问题转化为一系列无约束极值问题来求解, 故此方法也称为序列无约束最小化方法. 在无约束问题的求解过程中, 对企图违反约束的那些点给出相应的惩罚约束, 迫使这一系列的无约束问题的极小点不断地向可行域靠近 (若在可行域外部), 或者一直在可行域内移动 (若在可行域内部), 直到收敛到原问题的最优解为止.

常用的制约函数可分为两类: 惩罚函数 (简称罚函数) 和障碍函数, 从方法上讲分为外点法 (或外部惩罚函数法) 和内点法 (或内部惩罚函数法, 即障碍函数法).

外点法: 对违反约束条件的点在目标函数中加入相应的“惩罚约束”, 而对可行点不予惩罚, 此方法的迭代点一般在可行域的外部移动.

内点法: 对企图从内部穿越可行域边界的点在目标函数中加入相应的“障碍约束”, 距边界越近, 障碍越大, 在边界上给以无穷大的障碍, 从而保证迭代一直在可行域内部进行.

12.4.1 外点法(惩罚函数法)

(1) 对于等式约束的问题

$$\begin{cases} \min f(X), \\ h_i(X) = 0, i = 1, 2, \dots, m, \end{cases} \quad (12.6)$$

作辅助函数

$$P_1(X, M) = f(X) + M \sum_{j=1}^m h_j^2(X).$$

取 M 为充分大的正数, 则问题 (12.6) 可以转化为求无约束问题 $\min_x P_1(X, M)$ 的解的问题. 如果其最优解 X^* 满足或近似满足 $h_j(X^*) = 0$ ($j = 1, 2, \dots, m$), 即是原问题 (12.6) 的可行解, 或近似可行解, 则 X^* 就是原问题 (12.6) 的最优解或近似解.

由于 M 是充分大的数, 在求解的过程中使对求 $\min_x P_1(X, M)$ 起着限制作用, 即限制 X^* 成为极小点, 因此, 称 $P_1(X, M)$ 为惩罚函数, 其中第二项 $M \sum_{j=1}^m h_j^2(X)$ 称为惩罚项, M 称为惩罚因子.

(2) 对于不等式约束的问题 (12.3), 同样可构造惩罚函数, 即对充分大的 M 作:

$$P_2(X, M) = f(X) + M \sum_{j=1}^m [\min\{0, g_j(X)\}]^2$$

则问题 (12.3) 可以转化为求 $\min_x P_2(X, M)$ 的问题, 其解之间的关系同 (1) 类似.

(3) 对于一般的问题 (12.2) 也可构造出惩罚函数, 即对于充分大的 M 作辅助函数:

$$P_3(X, M) = f(X) + MP(X)$$

其中 $P(X) = \sum_{i=1}^m |h_i(X)|^2 + \sum_{j=1}^l [\min\{0, g_j(X)\}]^2$, 则将原问题 (12.2) 可化为求解 $\min_x P_3(X, M)$ 的问题.

在实际中, 惩罚因子 M 的选择十分重要, 一般的策略是取一个趋向无穷大的严格递增正数列 $\{M_k\}$, 逐个求解

$$\min_x P_3(X, M_k) = f(X) + M_k P(X).$$

于是可得到一个极小点的序列 $\{X_k^*\}$, 在一定的条件下, 这个序列收敛于原问题的最优解. 因此, 这种方法又称为序列无约束极小化方法, 简称为 SUMT 方法.

SUMT 方法的迭代步骤如下:

1) 取 $M_1 > 0$, (例如 $M_1 = 1$) 允许误差 $\varepsilon > 0$, 并取 $k = 1$;

2) 以 $X^{(k-1)}$ 为初始值, 求解无约束问题

$$\min_X P_i(X, M_k) = f(X) + M_k P_i(X), \quad 1 \leq i \leq 3,$$

其中

$$P_1(X) = \sum_{j=1}^m h_j^2(X), \quad P_2(X) = \sum_{j=1}^m [\min\{0, g_j(X)\}]^2,$$

$$P_3(X) = \sum_{i=1}^m |h_i(X)|^2 + \sum_{j=1}^l [\min\{0, g_j(X)\}]^2;$$

3) 若 $M_k P_i(X^{(k)}) < \varepsilon$, 则停止计算, 即得到近似解 $X^{(k)}$; 否则令 $M_{k+1} = cM_k$ (例如 $c=5$ 或 10), 令 $k = k+1$, 转回 2)。

12.4.2 内点法(障碍函数法)

由于内点法总是在可行域内进行的, 并一直保持在可行域内进行搜索, 因此这种方法只适用于不等式约束的问题(12.3)。作辅助函数(障碍函数)

$$Q(X, r) = f(X) + rB(X),$$

其中 $B(X)$ 是连续函数, $rB(X)$ 称为障碍项; r 为充分小的正数, 称之为障碍因子。

注意到, 当点 X 趋向于 R 的边界时, 要使 $B(X)$ 趋向于正无穷大, 则 $B(X)$ 的最常用的两种形式为

$$B(X) = \sum_{j=1}^m \frac{1}{g_j(X)} \quad \text{和} \quad B(X) = - \sum_{j=1}^m \ln[g_j(X)].$$

由于 $B(X)$ 的存在, 在 R 的边界上形成了“围墙”, 对迭代点的向外移动起到了阻挡作用, 而越靠近边界阻力就越大。这样, 当点 X 趋向于 R 的边界时, 则障碍函数 $Q(X, r)$ 趋向于正无穷大; 否则, $Q(X, r) \approx f(X)$ 。因此, 问题可以转化为求解问题:

$$\min_{X \in R_0} Q(X, r)$$

其中 $R_0 = \{X | g_j(X) > 0, j=1, 2, \dots, m\}$ 表示 R 的内部。

根据 $Q(X, r)$ 定义, 显然障碍因子越小, $\min_{X \in R_0} Q(X, r)$ 的解就越接近于原问题的解, 因此, 在实际计算中, 也采用 SUMT 方法, 即取一个严格单调减少且趋于零的障碍因子数列 $\{r_k\}$, 对于每一个 r_k , 从 R_0 内的某点出发, 求解 $\min_{X \in R_0} Q(X, r_k)$ 。

内点法的计算步骤如下:

(1) 取 $r_1 > 0$, (例如 $r_1 = 1$) 允许误差 $\varepsilon > 0$, 并取 $k = 1$;

(2) 以 $X^{(k-1)} \in R_0$ 为初始值, 求解无约束问题

$$\min_{X \in R_0} Q(X, r_k) = f(X) + r_k B(X),$$

不妨设极小点为 $X^{(k)}$;

(3) 若 $r_k B(X^{(k)}) < \varepsilon$, 则停止计算, 即得到近似解 $X^{(k)}$; 否则令 $r_{k+1} = \beta r_k$ (例如 $\beta = 1/5$ 或 $1/10$ 称为缩小系数), 令 $k = k + 1$, 转回(2).

12.5 奶制品的加工计划问题^[4]

12.5.1 问题的提出

一奶制品加工厂用牛奶生产 A_1, A_2 两种初级奶制品, 它们可以直接出售, 也可以分别加工成 B_1, B_2 两种高级奶制品再出售. 按目前技术每桶牛奶可加工成 2 kg A_1 和 3 kg A_2 , 每桶牛奶的买入价为 10 元, 加工费为 5 元, 加工时间为 15 小时. 每 kg A_1 可深加工成 0.8 kg B_1 , 加工费为 4 元, 加工时间为 12 小时; 每 kg A_2 可深加工成 0.7 kg B_2 , 加工费为 3 元, 加工时间为 10 小时. 初级奶制品 A_1, A_2 的售价分别为每 kg 10 元和 9 元, 高级奶制品 B_1, B_2 的售价分别为每 kg 30 元和 20 元. 工厂现有的加工能力每周总共 2 000 小时. 根据市场状况, 高级奶制品的需求量占全部奶制品需求量的 20% 至 40%. 试在供需平衡的条件下为该厂制订(一周的)生产计划, 使利润最大, 并进一步研究如下问题:

(1) 工厂拟拨一笔资金用于技术革新, 据估计可实现下列革新中的某一项: 总加工能力提高 10%; 各项加工费用均减少 10%; 初级奶制品 A_1, A_2 的产量提高 10%; 高级奶制品 B_1, B_2 的产量提高 10%. 问将资金用于哪一项革新, 这笔资金的上限(对于一周)应为多少;

(2) 该厂的技术人员又提出一项技术革新, 将原来的每桶牛奶可加工成 2 kg A_1 和 3 kg A_2 变为每桶牛奶可加工成 4 kg A_1 或 6 kg A_2 . 假设其他条件都不变, 问应否采用这项革新, 若采用, 生产计划如何;

(3) 根据市场经济规律, 初级奶制品 A_1, A_2 的售价都要随着它们二者销售量的增加而减少, 同时, 在深加工过程中, 单位成本会随着它们各自加工数量的增加而减小. 在高级奶制品的需求量占全部奶制品需求量 20% 的情况下, 市场调查得到一批数据(如表 12-1). 试根据此市场实际情况对该厂的生产计划进行修订(设其他条件不变).

12.5.2 问题的分析

这是一个对实际生产计划经过简化的加工方案优化设计问题, 主要可以利用线性规划的方法来研究. 首先将题目给出的奶制品的加工和销售的过程用图

12-1 来描述.

表 12-1

| | | | | | | | | | | |
|------------|------|------|------|------|------|-----|------|------|-----|-----|
| A_1 售量 | 20 | 25 | 50 | 55 | 65 | 65 | 80 | 70 | 85 | 90 |
| A_2 售量 | 210 | 230 | 170 | 190 | 175 | 210 | 150 | 190 | 190 | 190 |
| A_1 售价 | 15.2 | 14.4 | 14.2 | 12.7 | 12.2 | 11 | 11.9 | 11.5 | 10 | 9.6 |
| A_2 售价 | 11 | 9.6 | 13 | 10.8 | 11.5 | 8.5 | 13 | 10 | 9.2 | 9.1 |
| A_1 深加工量 | 40 | 50 | 60 | 65 | 70 | 75 | 80 | 85 | 90 | 100 |
| A_1 深加工费 | 5.2 | 4.5 | 4.0 | 3.9 | 3.6 | 3.6 | 3.5 | 3.5 | 3.3 | 3.2 |
| A_2 深加工量 | 60 | 70 | 80 | 90 | 95 | 100 | 105 | 110 | 115 | 120 |
| A_2 深加工费 | 3.8 | 3.3 | 3.0 | 2.9 | 2.9 | 2.8 | 2.8 | 2.8 | 2.7 | 2.7 |

注：表中数量单位是 kg, 费用单位是元/kg

由题意可知： A_1, B_1, A_2, B_2 的售价分别为 $p_1 = 10, p_2 = 30, p_3 = 9, p_4 = 20$ (元/kg). 牛奶的购入和加工费用为 $q_1 = 10 + 5 = 15$ (元/桶), 深加工 A_1, A_2 的费用分别为 $q_2 = 4, q_3 = 3$ (元/kg). 每桶牛奶可加工成 $a = 2$ kg A_1 和 $b = 3$ kg A_2 , 每千克 A_1 可深加工成 $c = 0.8$ kg B_1 , 每千克 A_2 可深加工成 $d = 0.7$ kg B_2 , 每桶牛奶的加工时间为 15 h, 每千克 A_1, A_2 的深加工时间分别为 12, 10 (h), 工厂的总加工能力为 $S = 2\,000$ h. B_1, B_2 的销售量 (即产量) 占全部奶制品的比例为 20% 至 40%.

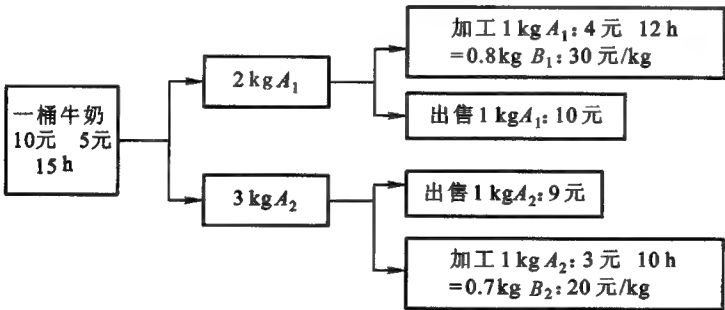


图 12-1 奶制品的加工和销售

记出售 A_1, B_1 的数量分别为 x_1, x_2 (kg), 出售 A_2, B_2 的数量分别为 x_3, x_4 (kg), 生产的 A_1, A_2 的数量分别为 x_5, x_6 (kg), 购入和加工牛奶的数量为 x_7 桶, 深加工的 A_1, A_2 的数量分别为 x_8, x_9 (kg).

12.5.3 模型的建立与求解

根据上面的分析, 在供需平衡的条件下, 使得利润最大的生产计划应满足下面的线性规划模型

$$\begin{aligned}
 \max z &= 10x_1 + 30x_2 + 9x_3 + 20x_4 - 15x_7 - 4x_8 - 3x_9, \\
 \text{s. t. } &\begin{cases} x_5 = 2x_7, x_6 = 3x_7, x_2 = 0.8x_8, x_4 = 0.7x_9, \\ x_5 = x_1 + x_8, x_6 = x_3 + x_9, \\ 15x_7 + 12x_8 + 10x_9 \leq 2000, \\ 0.2(x_1 + x_2 + x_3 + x_4) \leq x_2 + x_4 \leq 0.4(x_1 + x_2 + x_3 + x_4), \\ x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9 \geq 0. \end{cases}
 \end{aligned} \tag{12.7}$$

利用 MATLAB 求解,并作拉格朗日 (Lagrange) 分析可得

$$\begin{aligned}
 X &= (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9) \\
 &= (55.184\ 6, 65.040\ 7, 204.878\ 0, 0, 136.585\ 4, 204.878\ 0, \\
 &\quad 68.292\ 7, 81.300\ 8, 0), \\
 \text{Lag} &= (8.097\ 6, 7.097\ 6, -37.609\ 8, -35.842\ 0, 8.097\ 6, 7.097\ 6, 1.499\ 2, \\
 &\quad 9.512\ 2, 0, 0, 0, 0, 8.232\ 3, 0, 0, 0, 0), z = 2\ 998.4.
 \end{aligned}$$

对所解得的 X 值作适当的取整处理可以得到(一周的)生产计划为:购入、加工 68 桶牛奶,加工成 136 kg A_1 ,及 204 kg A_2 ,其中 55 kg A_1 直接出售,81 kg A_1 再加工成 4.8 kg B_1 出售,而 204 kg A_2 则全部直接出售,这样可获得利润为 2 986 元。

由 Lag 值可知,加工能力 2 000 h 已用足,且每增加 1 h 可获利 1.499 2 元;高级奶制品的产量占全部奶制品产量达到下限 20%。而按上面给出的计划实施可算出加工能力为 1 992 h,高级奶制品的产量比例为 20.01%,因此,此计划是可行的。

如果在建模时就要求购入和加工牛奶的桶数 x_7 为整数,那么线性规划模型 (12.7) 将变为混合整数规划模型,可用 LINDO 软件求解得

$$\begin{aligned}
 X &= (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9) \\
 &= (54.333\ 3, 65.333\ 3, 204, 0, 136, 204, 68, 81.666\ 7, 0),
 \end{aligned}$$

而且总获利润为 $z = 2\ 992.7$ 。与上面的结果稍有差别。

12.5.4 关于革新项目的研究

问题(1): 合理使用革新资金

(1) 总加工能力提高 10%, 即 $S = 2\ 200$ 小时, 由 (12.7) 式求解得最大利润为 $z = 3\ 298.2$ 元。

(2) 各项加工费用均减少 10%, 即 $q_1 = 14.5$ 元/桶, $q_2 = 3.6$, $q_3 = 2.7$ (元/kg), 由 (10.7) 式得最大利润为 $z = 3\ 065$ 元。

(3) 初级奶制品 A_1, A_2 的产量提高 10%, 即 $a = 2.2, b = 3.3$ (kg), 由 (10.7)

式得最大利润为 $z = 3\,242.5$ 元.

(4) 高级奶制品 B_1, B_2 的产量提高 10%, 即 $c = 0.88, d = 0.77$ (kg), 由 (12.7) 式得最大利润为 $z = 3\,233.8$ 元.

比较以上四项革新项目所得的利润可知, 应将资金用于提高加工能力上, 一周最大获利为 3 298.2 元, 比原获利增加 $3\,298.2 - 2\,998.4 = 299.8$, 所以这笔资金的上限 (对于一周) 应为 300 元. 实际上, 这个结果也可由 $\text{Lag}(7) \times 200 = 1.499\,2 \times 200$ 得到.

问题(2): 论证新的革新方案

题目给出的又一技术革新, 是将原来的每桶牛奶可加工成品 2 kg A_1 和 3 kg A_2 变为每桶牛奶可加工成 4 kg A_1 或 6 kg A_2 . 只要将模型 (12.7) 中的约束

条件 $x_5 = 2x_7, x_6 = 3x_7$ 改为 $\frac{x_5}{4} + \frac{x_6}{6} = x_7$, 利用 MATLAB 求解得

$$\begin{aligned} X &= (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9) \\ &= (0, 67.368\,4, 269.473\,7, 0, 84.210\,5, 269.473\,7, 65.964\,9, 84.210\,5, 0) \end{aligned}$$

对 X 作适当的取整处理得到相应的生产计划为: 购入、加工 66 桶牛奶, 用 21 桶加工成 84 kg A_1 , 用 45 桶加工成 270 kg A_2 , 84 kg A_1 全部再加工成 67.2 kg B_1 出售, 而 270 kg A_2 则全部直接出售, 这样总获利仍为 3 120 元, 大于原来的 2 986 元, 加工时间为 1 998 h, 高级奶制品的产量比例为 19.93%. 因此应该采用这项技术革新. 这是由于每桶牛奶可加工成 4 kg A_1 或 6 kg A_2 , 与原来的每桶牛奶可加工成品 2 kg A_1 和 3 kg A_2 相比, 虽然看起来 A_1, A_2 的基本产量未变, 但是生产的安排更加灵活

问题(3): 修订生产计划

根据市场规律和题意, A_1, A_2 的销售价格都要随着二者销售量的增加而减少, 为此, 设 A_1, A_2 的单价 $p_1(x_1, x_3), p_3(x_1, x_3)$ 均是 x_1, x_3 的减函数, 且由所给数据作如下线性函数拟合:

$$p_1(x_1, x_3) = a_1 + b_1x_1 + c_1x_3, p_3(x_1, x_3) = a_2 + b_2x_1 + c_2x_3. \quad (12.8)$$

类似地, A_1, A_2 深加工的单位成本均随着它们各自的产量的增加而减少, 为此, 设 A_1, A_2 深加工的单位成本 $q_2(x_8), q_3(x_9)$ 均是减函数, 且根据所给数据作离散点的图形容易看出, 可应用如下的二次函数拟合:

$$q_2(x_8) = e_1x_8^2 + f_1x_8 + g_1, \quad q_3(x_9) = e_2x_9^2 + f_2x_9 + g_2. \quad (12.9)$$

作最小二乘拟合和统计分析可得到

$$a_1 = 24.729\,9, \quad b_1 = -0.093\,7, \quad c_1 = -0.035\,6, \quad R^2 = 0.993\,3;$$

$$a_2 = 29.957\,5, \quad b_2 = -0.056\,3, \quad c_2 = -0.083\,9, \quad R^2 = 0.987\,3;$$

$$e_1 = 0.000\,553, \quad f_1 = -0.108\,4, \quad g_1 = 8.587\,9, \quad R^2 = 0.984\,9;$$

$$e_2 = 0.000\,368, \quad f_2 = -0.082\,2, \quad g_2 = 7.327\,2, \quad R^2 = 0.962\,6.$$

由相关系数 R 值可以看出,拟合结果是令人满意的.

将拟合结果(12.8)和(12.9)式代入(12.7)的目标函数中,并对约束条件做相应的修改,可以得到一个非线性规划模型

$$\begin{aligned} \max z = & (a_1 + b_1x_1 + c_1x_3)x_1 + 30x_2 + (a_2 + b_2x_1 + c_2x_3)x_3 + 20x_4 \\ & - 15x_7 - (e_1x_8^2 + f_1x_8 + g_1)x_8 - (e_2x_9^2 + f_2x_9 + g_2)x_9, \quad (12.10) \\ \text{s. t. } & \begin{cases} x_5 = 2x_7, x_6 = 3x_7, x_2 = 0.8x_8, x_4 = 0.7x_9, \\ x_5 = x_1 + x_8, x_6 = x_3 + x_9, \\ 15x_7 + 12x_8 + 10x_9 \leq 2000, \\ 0.2(x_1 + x_2 + x_3 + x_4) = x_2 + x_4, \\ x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9 \geq 0. \end{cases} \end{aligned}$$

将拟合系数代入(12.10)式中,利用 LINGO 软件求得

$$\begin{aligned} X = & (x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9) \\ = & (47.35384, 55.71040, 175.4878, 0, 116.9918, 175.4878, 58.49592, \\ & 69.6380, 0), \\ z = & 3405.406. \end{aligned}$$

对 X 的值作适当的取整处理,可以得到一周的生产计划的修订方案:购入、加工 58 桶牛奶,加工成 116 kg A_1 , 及 174 kg A_2 , 其中 47 kg A_1 直接销售, 69 kg A_1 再加工成 55 kg B_1 出售, 而 174 kg A_2 全部直接出售, 这样可总获利为 3 398 元, 并且可算得加工时间为 1 698 h, 高级奶制品的产量比例为 19.93%.

与原方案比较,购入牛奶桶数、加工时间均减少,获利反而增加. 其原因是根据市场规律和所给数据,采用这个新的销量和加工量,使 A_1 的售价 p_1 由 10 增为 14.045 48, A_2 的售价 p_2 由 9 增为 12.568 05, 而 A_1 的深加工费用 q_2 由 4 降为 3.720 887, 显然这一方案要优于原来的方案.

12.6 参考案例与参考文献

1. 参考案例

- (1) 飞行管理问题——文献[3]:84—109
- (2) 节水洗衣机问题——文献[3]:190—219
- (3) 投资的收益和风险问题——文献[3]:327—368
- (4) 钻井布局问题——文献[3]:441—476
- (5) 钢管订购和运输问题——文献[3]:538—573
- (6) 彩票中的数学问题——文献[10]:107—116

2. 参考文献

- [1] 姜启源. 数学模型. 第三版. 北京:高等教育出版社, 2003

- [2] 赵静,但琦等. 数学建模与数学实验. 北京:高等教育出版社,2000
- [3] 全国大学生数学建模竞赛组委会. 全国大学生数学建模竞赛优秀论文汇编. 北京:中国物价出版社,2002
- [4] 韩中庚,姜启源. 奶制品加工计划的设计模型. 信息工程大学学报, 2002,3(1):40—42
- [5] 汪国强. 数学建模优秀案例选编. 广州:华南理工大学出版社,1998
- [6] 杨超学. 运筹学. 北京:科学出版社,2004
- [7] 何坚勇. 运筹学基础. 北京:清华大学出版社,1999
- [8] 施光燕等. 最优化方法. 北京:高等教育出版社,1999
- [9] 《运筹学》教材编写组. 运筹学. 修订版. 北京:清华大学出版社,1990
- [10] 韩中庚. “彩票中的数学”问题的优化模型与评述. 工程数学学报, 2003;20(5),107—116

第十三章 动态规划方法

动态规划是一种将复杂的多阶段决策问题转化为一系列比较简单的最优化问题的方法,它的基本特征是优化过程的多阶段性.

动态规划是求解某一类问题的一种方法,是考查问题的一种途径,而不是一种算法,它没有标准的数学表达式和具体的算法,必须具体问题具体分析处理.

动态规划是一种用于处理多阶段决策问题的数学方法,主要是先将一个复杂的问题分解成相互联系的若干阶段,每个阶段即为一个子问题,然后逐个解决,当每个阶段的决策确定之后,整个过程的决策也就确定了.阶段一般用时间段来表示(即与时间有关),这就是“动态”的含义,把这种处理问题的方法称为动态规划方法.

13.1 动态规划的基本概念和基本方程

13.1.1 动态规划的基本概念

(1) 阶段和阶段变量:阶段是指一个问题需要做出决策的步骤,即把问题的过程分为若干个相互联系的阶段,使能按阶段的次序求解,描述阶段的变量称为阶段变量,常用 k 表示.

(2) 状态和状态变量:在多阶段决策过程中,每一阶段都具有一些特征(自然状况或客观条件),这就是状态,用来描述状态的变量称为状态变量.通常第 k 阶段的状态变量用 $s_k (k=1, 2, \dots, n)$ 表示,它的取值可以是一个数、一组数或一个向量等,状态变量可取值的全体所构成的集合称为可达状态集合(或允许状态集合),用 $S_k (k=1, 2, \dots, n)$ 表示.

(3) 决策和决策变量:当过程处于某一阶段的某个状态时,可以做出不同的决定(或选择),从而确定下一阶段的状态,这种决定称为决策,描述决策的变量称为决策变量,用 $x(s_k)$ 表示第 k 阶段 $s_k (k=1, 2, \dots, n)$ 状态的决策变量.决策变量的取值范围称为允许决策集合,用 $D_k(s_k)$ 表示第 k 阶段状态 $s_k (k=1, 2, \dots, n)$ 的允许决策集合,即 $x_k(s_k) \in D_k(s_k) (k=1, 2, \dots, n)$.

(4) 策略与子策略:一个按顺序排列的决策组成的集合称为策略.由第 k

阶段开始到终止状态为止的过程,称为问题的后部子过程,或 k 子过程. 由 k 子过程的每一阶段的决策按顺序排列组成的决策函数序列 $\{x_k(s_k), \dots, x_n(s_n)\}$, 称为 k 子策略, 记为 $p_{k,n}(s_k)$, 即

$$p_{k,n}(s_k) = \{x_k(s_k), x_{k+1}(s_{k+1}), \dots, x_n(s_n)\}, k = 1, 2, \dots, n.$$

当 $k=1$ 时, 此决策函数序列称为全过程的一个策略, 记为 $p_{1,n}(s_1)$, 即

$$p_{1,n}(s_1) = \{x_1(s_1), x_2(s_2), \dots, x_n(s_n)\}.$$

可供选择的策略范围称为允许策略集合, 用 P 表示, 从允许策略集合中找出达到最优效果的策略称为最优策略.

(5) 状态转移函数: 状态转移函数是在确定多阶段决策过程中, 由一个状态到另一个状态的演变过程.

如果给定了第 k 阶段状态变量 s_k 和该阶段的决策变量 $x_k(s_k)$, 则第 $k+1$ 阶段的状态变量 s_{k+1} 的值也随之而定, 即 s_{k+1} 随 s_k 和 $x_k(s_k)$ 的变化而变化. 这种对应关系记为 $s_{k+1} = T_k(s_k, x_k(s_k))$, 称为状态转移方程, $T_k(s_k, x_k)$ 称为状态转移函数.

(6) 指标函数(回收函数): 在多阶段决策过程中, 用来衡量所实现过程优劣的一种数量指标, 称为指标函数, 或回收函数. 它是定义在全过程或所有子过程上的数量函数, 即是各阶段的状态和决策变量的函数, 记为 $V_{k,n}$, 即

$$V_{k,n} = V_{k,n}(s_k, x_k, s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1}), k = 1, 2, \dots, n.$$

指标函数具有可分离性和递推关系:

$$\begin{aligned} & V_{k,n}(s_k, x_k, s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1}) \\ &= \varphi_k[s_k, x_k, V_{k+1,n}(s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1})]. \end{aligned}$$

特别地, 有两种常见的形式:

1) 全过程和任一子过程的指标函数是它所包含的各阶段指标函数的和, 即

$$V_{k,n}(s_k, x_k, s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1}) = \sum_{j=k}^n v_j(s_j, x_j).$$

递推关系为

$$\begin{aligned} & V_{k,n}(s_k, x_k, s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1}) \\ &= v_k(s_k, x_k) + V_{k+1,n}(s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1}). \end{aligned}$$

2) 全过程和任一子过程的指标函数是它所包含的各阶段指标函数的乘积, 即

$$V_{k,n}(s_k, x_k, s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1}) = \prod_{j=k}^n v_j(s_j, x_j).$$

递推关系为

$$\begin{aligned} & V_{k,n}(s_k, x_k, s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1}) = \\ & v_k(s_k, x_k) \cdot V_{k+1,n}(s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1}). \end{aligned}$$

(7) 最优值函数: 从第 k 阶段的状态 s_k 开始到第 n 阶段的终止状态 s_{n+1} 的

过程, 采取最优策略所得到的指标函数值, 称为**最优值函数**, 记为 $f_k(s_k)$ ($k=1, 2, \dots, n$), 即

$$f_k(s_k) = \underset{\{x_k, s_{k+1}, \dots, s_n\}}{\text{opt}} V_{k,n}(s_k, x_k, s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1}).$$

说明: 在实际中, 指标函数的含义可以是距离、利润、成本、时间、产品的产量、资源消耗等.

13.1.2 动态规划的基本条件

要对一个实际问题建立动态规划模型必须要满足下面的五个要素:

- (1) 问题可以化为恰当的 n 个阶段;
- (2) 正确选择状态变量 s_k , 使它既能表达过程, 又要具有无后效性和可知性:

1) **无后效性**: 如果某阶段状态已给定, 则以后过程的发展不受以前各阶段状态的影响, 也就是说当前状态就是未来过程的初始状态;

2) **可知性**: 规定的各阶段状态变量的值, 由直接或间接地都是可以知道的;

(3) 确定决策变量 x_k 及每一阶段的允许决策集合 $D_k(s_k)$;

(4) 写出状态转移方程: $s_{k+1} = T_k(s_k, x_k)$, $k=1, 2, \dots, n$;

(5) 正确写出指标函数 $V_{k,n}$ 的关系, 它满足下列性质:

1) 它是过程各阶段状态变量和决策变量的函数;

2) 具有可分离性和递推关系, 即

$$\begin{aligned} & V_{k,n}(s_k, x_k, s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1}) \\ &= \varphi_k[s_k, x_k, V_{k+1,n}(s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1})] \end{aligned}$$

3) 函数 $\varphi_k[s_k, x_k, V_{k+1,n}(s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1})]$ 是关于 $V_{k+1,n}$ 严格单调的. 现在的问题是: 对于动态规划的问题如何求解呢?

13.1.3 动态规划的基本方程

1. 逆序解法

设指标函数的形式为

$$V_{k,n}(s_k, x_k, s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1}) = \sum_{j=k}^n v_j(s_j, x_j),$$

且具有上面的三条性质, 则

$$V_{k,n} = v_k(s_k, x_k) + V_{k+1,n}(s_{k+1}, x_{k+1}, \dots, s_n, x_n, s_{n+1}).$$

如果初始状态 s_k 给定, 则决策变量 $x_k(s_k)$ 随之确定, k 子过程的策略 $p_{k,n}(s_k)$ 也就确定, 从而指标函数 $V_{k,n}$ 也同时确定了. 于是, 指标函数可以看成是初始状态和策略的函数, 即对 k 子过程的指标函数为 $V_{k,n}[s_k, p_{k,n}(s_k)]$, 且有递

推关系

$$V_{k,n}[s_k, p_{k,n}(s_k)] = v_k(s_k, x_k) + V_{k+1,n}[s_{k+1}, p_{k+1,n}(s_{k+1})]$$

其中子策略为 $p_{k,n}(s_k) = \{x_k(s_k), p_{k+1,n}(s_{k+1})\}$, 即为决策变量 $x_k(s_k)$ 和子策略 $p_{k+1,n}(s_{k+1})$ 的集合.

如果用 $p_{k,n}^*(s_k)$ 表示以第 k 阶段状态 s_k 为初始状态的后部子过程所有子策略中的最优子策略, 则最优值函数为

$$f_k(s_k) = V_{k,n}[s_k, p_{k,n}^*(s_k)] = \underset{p_{k,n}}{\text{opt}} V_{k,n}[s_k, p_{k,n}(s_k)],$$

其中

$$\begin{aligned} \underset{p_{k,n}}{\text{opt}} V_{k,n}[s_k, p_{k,n}(s_k)] &= \underset{\{x_k, p_{k+1,n}\}}{\text{opt}} \{v_k(s_k, x_k) + V_{k+1,n}[s_{k+1}, p_{k+1,n}(s_{k+1})]\} \\ &= \underset{x_k \in D_k(s_k)}{\text{opt}} \{v_k(s_k, x_k) + \underset{p_{k+1,n}}{\text{opt}} V_{k+1,n}[s_{k+1}, p_{k+1,n}(s_{k+1})]\}. \end{aligned}$$

而

$$f_{k+1}(s_{k+1}) = \underset{p_{k+1,n}}{\text{opt}} V_{k+1,n}[s_{k+1}, p_{k+1,n}(s_{k+1})],$$

故

$$\begin{cases} f_k(s_k) = \underset{x_k \in D_k(s_k)}{\text{opt}} \{v_k(s_k, x_k) + f_{k+1}(s_{k+1})\}, & k = n, n-1, \dots, 2, 1, \\ \text{边界条件为 } f_{n+1}(s_{n+1}) = 0 \text{ 或 } f_n(s_n) = v_n(s_n, x_n). \end{cases}$$

(13.1)

此为动态规划逆序解法的基本方程.

求解时, 由边界条件, 从 $k=n$ 开始, 由后向前逆推, 逐段求最优决策和过程的最优值, 最后求出 $f_1(s_1)$ 为问题的最优解.

2. 顺序解法

设过程的第 k 阶段的状态为 s_k , 其决策变量 x_k 表示当状态处于 s_{k+1} 的决策, 即由 $x_k(s_{k+1})$ 确定, 则状态转移方程为 $s_k = T_k(s_{k+1}, x_k)$, k 阶段的允许决策集合记为 $D'_k(s_{k+1})$, 指标函数定义为 $V_k(s_{k+1}, x_k, s_k, x_{k-1}, \dots, x_1, s_1)$, 其最优值函数为

$$f_k(s_{k+1}) = \underset{x_k \in D'_k(s_{k+1})}{\text{opt}} \{V_k(s_{k+1}, x_k, \dots, x_1, s_1)\},$$

则

$$\begin{cases} f_k(s_{k+1}) = \underset{x_k \in D'_k(s_{k+1})}{\text{opt}} \{v_k(s_{k+1}, x_k) + f_{k-1}(s_k)\}, & k = 1, 2, \dots, n, \\ \text{边界条件为 } f_0(s_1) = 0. \end{cases} \quad (13.2)$$

此为动态规划顺序解法的基本方程.

求解时, 由边界条件, 从 $k=1$ 开始, 由前向后顺推, 逐段求出最优决策和过

程的最优值,最后求出 $f_n(s_{n+1})$ 为问题的最优解.

13.2 动态规划的求解方法

动态规划的求解方法有逆序解法和顺序解法两种:

如果已知过程的初始状态 s_1 , 则用逆序解法; 如果已知过程的终止状态 s_{n+1} , 则用顺序解法.

13.2.1 逆序解法

设已知初始状态为 s_1 , 用 $f_k(s_k)$ 表示从第 k 阶段初始状态 s_k 到第 n 阶段的最优值.

第 n 阶段: 指标函数的最优值记为 $f_n(s_n) = \underset{x_n \in D_n(s_n)}{\text{opt}} v_n(s_n, x_n)$, 此为一维极值问题, 不妨设有最优解 $x_n = x_n(s_n)$, 于是可有最优值 $f_n(s_n)$.

第 $n-1$ 阶段: 类似地有

$$f_{n-1}(s_{n-1}) = \underset{x_{n-1} \in D_{n-1}(s_{n-1})}{\text{opt}} \{v_{n-1}(s_{n-1}, x_{n-1}) * f_n(s_n)\},$$

其中 $s_n = T_{n-1}(s_{n-1}, x_{n-1})$, 可解得最优解 $x_{n-1} = x_{n-1}(s_{n-1})$, 于是最优值为 $f_{n-1}(s_{n-1})$.

不妨设第 $k+1$ 阶段的最优解为 $x_{k+1} = x_{k+1}(s_{k+1})$ 和最优值 $f_{k+1}(s_{k+1})$, 则对于第 k 阶段有

$$f_k(s_k) = \underset{x_k \in D_k(s_k)}{\text{opt}} \{v_k(s_k, x_k) * f_{k+1}(s_{k+1})\},$$

其中 $*$ 表示“+”或“ \times ”, $s_{k+1} = T_k(s_k, x_k)$, 可解得最优解 $x_k = x_k(s_k)$ 和最优值为 $f_k(s_k)$.

依此类推, 直到第 1 阶段, 有 $f_1(s_1) = \underset{x_1 \in D_1(s_1)}{\text{opt}} \{v_1(s_1, x_1) * f_2(s_2)\}$, 其中 $s_2 = T_1(s_1, x_1)$, 可解得最优解 $x_1 = x_1(s_1)$ 和最优值为 $f_1(s_1)$.

由于已知 s_1 , 则可知 x_1 与 $f_1(s_1)$. 从而可知 $s_2, x_2, f_2(s_2)$, 按上面的过程反推回去, 即可得到每一阶段和全过程的最优决策.

13.2.2 顺序解法

设已知终止状态为 s_{n+1} , 用 $f_k(s_{k+1})$ 表示从第 1 阶段初始状态 s_1 到第 k 阶段末的结束状态 s_{k+1} 的最优值.

第一阶段: 指标函数的最优值记为 $f_1(s_2) = \underset{x_1 \in D_1(s_1)}{\text{opt}} v_1(s_1, x_1), s_1 = T_1(s_2, x_1)$, 可解得最优解 $x_1 = x_1(s_2)$ 和最优值 $f_1(s_2)$.

第二阶段: 类似地有

$$f_2(s_3) = \operatorname{opt}_{x_2 \in D_2(s_2)} \{v_2(s_2, x_2) * f_1(s_2)\},$$

其中 $s_2 = T_2(s_3, x_2)$, 可解得最优解 $x_2 = x_2(s_3)$, 于是最优值为 $f_2(s_3)$.

不妨设第 k 阶段有

$$f_k(s_{k+1}) = \operatorname{opt}_{x_k \in D_k(s_k)} \{v_k(s_k, x_k) * f_{k-1}(s_k)\},$$

其中 $s_k = T_k(s_{k+1}, x_k)$ 解得最优解为 $x_k = x_k(s_{k+1})$ 和最优值 $f_k(s_{k+1})$.

依此类推, 直到第 n 阶段有

$$f_n(s_{n+1}) = \operatorname{opt}_{x_n \in D_n(s_n)} \{v_n(s_n, x_n) * f_{n-1}(s_n)\},$$

其中 $s_n = T_n(s_{n+1}, x_n)$, 可解得最优解 $x_n = x_n(s_{n+1})$ 和最优值为 $f_n(s_{n+1})$.

由于已知 s_{n+1} , 则可知 x_n 与 $f_n(s_{n+1})$. 从而可知 $s_n, x_{n-1}, f_{n-1}(s_n)$, 按上面的过程反推回去, 直到 $s_1, x_1, f_1(s_2)$, 即得到整个过程和各阶段的最优决策.

13.3 动态规划方法的应用

13.3.1 一类静态规划的动态规划解法

所谓的静态规划是与时间概念无关的规划问题, 例如: 线性规划、目标规划、整数规划、非线性规划等. 而动态规划是与时间概念有关的规划问题, 动态规划的特点就是将问题按时间或空间特征而分成若干个阶段, 从而将整个决策过程化为多阶段的决策问题. 这也是动态规划与静态规划的区别. 对某些静态规划问题可以人为地引入时间因素, 视为一个按阶段进行的动态规划问题, 利用动态规划的方法求解.

在实际中, 许多问题都具有形式上相同的数学模型, 如载货问题、分配问题、背包问题等, 其模型的一般形式为

$$\begin{aligned} \max z &= \sum_{j=1}^n g_j(x_j) \left(\text{或} \prod_{j=1}^n g_j(x_j) \right), \\ \begin{cases} \sum_{j=1}^n a_j x_j \leq b, \\ 0 \leq x_j \leq c_j, j = 1, 2, \dots, n, \end{cases} \end{aligned} \quad (13.3)$$

其中 $g_j(x_j)$ 为已知函数, 当 $g_j(x_j)$ 均为线性函数时, 则是线性规划; 当 $g_j(x_j)$ 不全为线性函数时, 则是非线性规划; 当 x_j 为整数时, 则是整数规划.

一般解法: 把问题分为 n 个阶段, 取 x_k 为第 k 阶段的决策变量, 此时 $f_{n+1}(s_{n+1}) = 0$ 为边界条件, 则问题的基本方程为

$$\begin{cases} f_k(s_k) = \max [g_k(x_k) + f_{k+1}(s_{k+1})], \\ f_{n+1}(s_{n+1}) = 0. \end{cases}$$

状态变量为 $s_k = \sum_{i=k}^n a_i x_i \quad (k = 1, 2, \dots, n)$.

允许决策集合为 $D_k(s_k) = \left\{ x_k \mid 0 \leq x_k \leq \min\left(c_k, \frac{s_k}{a_k}\right) \right\}, k = 1, 2, \dots, n$.

允许状态集合为 $S_k = \{s_k \mid 0 \leq s_k \leq b\} \quad (1 < k \leq n), S_1 = \{b\}$.

状态转移函数为 $s_{k+1} = s_k - a_k x_k \quad (k = 1, 2, \dots, n-1), s_{n+1} = 0$.

注: 当决策变量 x_k 要求取整数时, 只要将允许决策集合限制在整数集合内取值即可.

13.3.2 资源分配问题的动态规划模型

设有某种资源总数量为 a , 用于生产 n 种产品, 如果分配数量 x_k 用于生产第 k 种产品, 其效益为 $g_k(x_k)$, 问如何分配资源使生产 n 种产品的总效益最大?

此问题的静态模型为

$$\begin{aligned} \max z &= \sum_{k=1}^n g_k(x_k), \\ \begin{cases} \sum_{k=1}^n x_k = a, \\ x_k \geq 0, \quad k = 1, 2, \dots, n. \end{cases} \end{aligned}$$

用动态规划方法求解, 构造动态规划模型如下:

设状态变量 s_k 表示分配用于生产第 k 种产品至第 n 种产品的资源数量, 决策变量 x_k 表示分配给生产第 k 种产品的资源数量, 状态转移方程为 $s_{k+1} = s_k - x_k (s_1 = a)$, 允许决策集合为 $D_k(s_k) = \{x_k \mid 0 \leq x_k \leq s_k\}, k = 1, 2, \dots, n$. 最优值函数 $f_k(s_k)$ 表示以 s_k 数量的资源分配给第 k 种产品至第 n 种产品所得的最大效益, 则问题的基本方程为

$$\begin{cases} f_k(s_k) = \max_{0 \leq x_k \leq s_k} \{g_k(x_k) + f_{k+1}(s_{k+1})\}, & k = n, n-1, \dots, 1, \\ f_n(s_n) = \max_{x_n = s_n} g_n(x_n), \text{ 或 } f_{n+1}(s_{n+1}) = 0, \end{cases}$$

其最优值为 $f_1(a)$.

13.3.3 生产与存贮问题的动态规划模型

设某企业(公司)对某种产品要制订一项 n 个阶段的生产(采购)计划. 已知初始库存量为 0, 每阶段的生产(或采购)的数量有上限为 m , 每阶段的需求量已知为 $d_k (k = 1, 2, \dots, n)$ (满足需要), 且在第 n 阶段结束时库存量为 0. 问如何制

订每个阶段的生产(采购)计划使总的成本费用最小?

设 x_k 为第 k 阶段的生产量(或采购量), s_k 为第 k 阶段结束时的库存量, 则 $s_k = \sum_{j=1}^k (x_j - d_j)$, $c_k(x_k)$ 表示第 k 阶段生产产品 x_k 时的成本费用, 包括准备成本 K 和产品成本 ax_k (a 为单位产品的成本), 即

$$c_k(x_k) = \begin{cases} 0, & x_k = 0, \\ K + ax_k, & x_k = 1, 2, \dots, m, \\ \infty, & x_k > m, \end{cases}$$

$h_k(s_k)$ 表示 k 阶段末有库存量为 s_k 所需的存贮费用, 则 k 阶段的费用为 $g_k = c_k(x_k) + h_k(s_k)$, 于是, 问题的静态模型为

$$\begin{aligned} \min g &= \sum_{k=1}^n [c_k(x_k) + h_k(s_k)], \\ \begin{cases} s_0 = 0, s_n = 0, \\ s_k = \sum_{j=1}^k (x_j - d_j) \geq 0, k = 1, 2, \dots, n-1, \\ 0 \leq x_k \leq m, x_k \text{ 为整数}, k = 1, 2, \dots, n, \end{cases} \end{aligned}$$

即为一整数非线性规划.

用动态规划方法求解, 采用顺序解法: s_k 为状态变量, x_k 为决策变量, 指标函数为 $g_k = c_k(x_k) + h_k(s_k)$, 状态转移方程为 $s_k = s_{k-1} + x_k - d_k$, 允许决策集合为 $D_k(s_k) = \{x_k | 0 \leq x_k \leq \sigma_k\}$, $\sigma_k = \min\{m, s_k + d_k\}$, 从第 1 阶段到第 k 阶段末的最小费用为 $f_k(s_k)$, 则问题的基本方程为

$$\begin{cases} f_k(s_k) = \min_{x_k \in D_k(s_k)} [c_k(x_k) + h_k(s_k) + f_{k-1}(s_{k-1})] \quad (k = 1, 2, \dots, n), \\ f_0(s_0) = 0, \text{ 或 } f_1(s_1) = \min_{0 \leq x_1 \leq \sigma_1} [c_1(x_1) + h_1(s_1)]. \end{cases}$$

从边界条件出发, 由基本方程可以算出每一阶段的 $f_k(s_k)$, 最后可得 $f_n(s_n) = f_n(0)$ 为全过程的最优值, 即最小费用.

13.3.4 背包问题的动态规划模型

设某人要装一个背包, 可装物品的质量限度为 a kg, 共有 n 种物品供选择 (即编号为 $1, 2, \dots, n$). 假设已知第 k 种物品质量为 w_k kg/件, 携带 x_k 件的使用价值为 $c_k(x_k)$ ($k = 1, 2, \dots, n$), 问应如何选择物品 (各多少件), 使总的使用价

值最大?

设 x_k 为选择第 k 种物品的件数, 则问题的静态规划模型为

$$\begin{aligned} \max f &= \sum_{k=1}^n c_k(x_k), \\ \begin{cases} \sum_{k=1}^n w_k x_k \leq a, \\ x_k \geq 0 \text{ 且为整数}, k = 1, 2, \dots, n. \end{cases} \end{aligned}$$

这是一个整数非线性规划问题.

问题的动态规划的模型: 设将 n 种物品分为 n 个阶段, 状态变量 s_k 表示选择第 1 种至第 k 种物品的总质量, 即 $s_k = \sum_{i=1}^k w_i x_i$, $s_{n+1} \leq a$, 决策变量为 x_k 表示选择第 k 种物品的件数, 则状态转移方程为 $s_k = s_{k+1} - x_{k+1} w_{k+1}$, 允许决策集合为 $D_k(s_k) = \left\{ x_k \mid 0 \leq x_k \leq \left\lfloor \frac{s_{k+1}}{w_k} \right\rfloor \right\}$, 最优值函数 $f_k(s_k)$ 表示总质量不超过 s_k kg 时,

选择第 1 种至第 k 种物品的最大使用价值, 即 $f_k(s_k) = \max_{\sum_{i=1}^k w_i x_i \leq s_k} \sum_{i=1}^k c_i(x_i)$. 于是可

得顺序解法的基本方程为

$$\begin{cases} f_k(s_{k+1}) = \max_{x_k \in D_k(s_k)} \{c_k(x_k) + f_{k-1}(s_k)\}, 2 \leq k \leq n, \\ f_1(s_2) = \max_{x_1 \in D_1(s_1)} \{c_1(x_1)\}. \end{cases}$$

用顺序解法求解得各阶段的最优值函数 $f_1(s_2), f_2(s_3), \dots, f_n(a)$ 以及相应的决策函数 $x_1(s_2), x_2(s_3), \dots, x_n(a), f_n(a)$ 为过程的最优值, 反推回去可得到最优策略.

13.3.5 复合系统工作可靠性问题的动态规划模型

设有 n 个部件组成的工作系统, 只要有一个部件失灵, 整个系统也就不能工作, 为了提高系统的可靠性, 在每个部件上均装有备用件, 并可自动替换. 实际上, 备用件越多整个系统正常工作的可靠性越大, 但系统的成本、质量、体积均相应增加, 工作精度会相应降低. 因此, 现在的问题是在上述的条件之下, 应如何选择各部件的备用件数, 使整个系统的工件可靠性最大?

设第 k 个部件装有 x_k 个备件, 正常工作的概率为 $p_k(x_k)$, 则整个系统正常工作的概率为 $p = \prod_{k=1}^n p_k(x_k)$ (衡量系统正常工作的可靠性的指标), 装一个部件

k 的备用件费用为 c_k , 质量为 w_k , 要求总费用不超过 c , 总质量不超过 w , 则问题的静态规划模型为

$$\begin{aligned} \max p &= \prod_{k=1}^n p_k(x_k), \\ &\begin{cases} \sum_{k=1}^n c_k x_k \leq c, \\ \sum_{k=1}^n w_k x_k \leq w, \\ x_k \geq 0, \text{ 为整数}, k = 1, 2, \dots, n. \end{cases} \end{aligned}$$

其问题可以转化为二维动态规划问题. 两个状态变量 s'_k, s''_k : 其 s'_k 表示由第 k 个到第 n 个部件所容许的总费用, s''_k 表示由第 k 个到第 n 个部件所容许的总质量, 决策变量 x_k 为部件 k 上装的备用件数, 状态转移方程为

$$\begin{cases} s'_{k+1} = s'_k - c_k x_k, \\ s''_{k+1} = s''_k - w_k x_k, \end{cases} \quad k = 1, 2, \dots, n.$$

允许决策集合为

$$D_k(s'_k, s''_k) = \{x_k \mid 0 \leq x_k \leq \min(\lfloor s'_k/c_k \rfloor, \lfloor s''_k/w_k \rfloor)\}.$$

最优值函数 $f_k(s'_k, s''_k)$ 表示从部件 k 到部件 n 的最大可靠性, 即

$$f_k(s'_k, s''_k) = \prod_{i=k}^n p_i(x_i),$$

于是可得问题的基本方程为

$$\begin{cases} f_k(s'_k, s''_k) = \max_{x_k \in D_k(s'_k, s''_k)} [p_k(x_k) \cdot f_{k+1}(s'_{k+1}, s''_{k+1})], \\ f_{n+1}(s'_{n+1}, s''_{n+1}) = 1, \quad k = n, n-1, \dots, 2, 1, \end{cases}$$

最后计算出 $f_1(c, w)$ 为系统的最大可靠性.

13.4 选拔队员与组队问题^[10]

13.4.1 问题的提出

在一年一度的美国 MCM 和全国大学生数学建模竞赛活动中, 任何一个参赛院校都会遇到如何选拔最优秀的队员和科学合理地组队问题, 这是一个最实际而且是首先需要解决的数学模型问题.

现假设有 20 名队员准备参加竞赛, 根据队员的能力和水平要选出 18 名优

秀队员分别组成 6 个队,每个队 3 名队员去参加比赛,选择队员主要考虑的条件依次为有关学科成绩(平均成绩)、智力水平(反映思维能力、分析问题和解决问题的能力等)、动手能力(计算机的使用和其他方面实际操作能力)、写作能力、外语水平、协作能力(团结协作能力)和其他特长、每个队员的基本条件量化后如表 13-1.

表 13-1 队员的基本条件

| 条件 数值 队员 | 学科 成绩 (I) | 智力 水平 (II) | 动手 能力 (III) | 写作 能力 (IV) | 外语 水平 (V) | 协作 能力 (VI) | 其他 特长 (VII) |
|----------------|-----------------|------------------|-------------------|------------------|-----------------|------------------|-------------------|
| A | 8.6 | 9.0 | 8.2 | 8.0 | 7.9 | 9.5 | 6 |
| B | 8.2 | 8.8 | 8.1 | 6.5 | 7.7 | 9.1 | 2 |
| C | 8.0 | 8.6 | 8.5 | 8.5 | 9.2 | 9.6 | 8 |
| D | 8.6 | 8.9 | 8.3 | 9.6 | 9.7 | 9.7 | 8 |
| E | 8.8 | 8.4 | 8.5 | 7.7 | 8.6 | 9.2 | 9 |
| F | 9.2 | 9.2 | 8.2 | 7.9 | 9.0 | 9.0 | 6 |
| G | 9.2 | 9.6 | 9.0 | 7.2 | 9.1 | 9.2 | 9 |
| H | 7.0 | 8.0 | 9.8 | 6.2 | 8.7 | 9.7 | 6 |
| I | 7.7 | 8.2 | 8.4 | 6.5 | 9.6 | 9.3 | 5 |
| J | 8.3 | 8.1 | 8.6 | 6.9 | 8.5 | 9.4 | 4 |
| K | 9.0 | 8.2 | 8.0 | 7.8 | 9.0 | 9.5 | 5 |
| L | 9.6 | 9.1 | 8.1 | 9.9 | 8.7 | 9.7 | 6 |
| M | 9.5 | 9.6 | 8.3 | 8.1 | 9.0 | 9.3 | 7 |
| N | 8.6 | 8.3 | 8.2 | 8.1 | 9.0 | 9.0 | 5 |
| O | 9.1 | 8.7 | 8.8 | 8.4 | 8.8 | 9.4 | 5 |
| P | 9.3 | 8.4 | 8.6 | 8.8 | 8.6 | 9.5 | 6 |
| Q | 8.4 | 8.0 | 9.4 | 9.2 | 8.4 | 9.1 | 7 |
| R | 8.7 | 8.3 | 9.2 | 9.1 | 8.7 | 9.2 | 8 |
| S | 7.8 | 8.1 | 9.6 | 7.6 | 9.0 | 9.6 | 9 |
| T | 9.0 | 8.8 | 9.5 | 7.9 | 7.7 | 9.0 | 6 |

假设所有队员接受了同样的培训,外部环境相同,竞赛中不考虑其他的随机因素的影响,竞赛水平的发挥只取决于表 13-1 中所给的各项条件,并且参赛队员都能正常发挥自己的水平. 现在的问题是

- (1) 在 20 名队员中选择 18 名优秀队员参加竞赛;
- (2) 确定一个最佳的组队使竞赛技术水平最高;
- (3) 给出由 18 名队员组成 6 个队的组队方案,使整体竞赛技术水平最高;并给出每个队的竞赛技术水平.

13.4.2 模型的假设

- (1) 问题中提供队员的基本条件充分地反映了每个队的真实能力和水平;
- (2) 每个队员的能力和水平在比赛中可以正常发挥,不受外界因素和环境的影响;
- (3) 同一个队三名队员的单项条件互不影响,而且具有互补性,即一个队的水平为最高者的水平;
- (4) 6 个队整体技术水平最高是在由问题(1)确定的最佳组队保持不变的条件整体技术水平最高.

13.4.3 模型的建立及求解

问题(1): 在 20 名队员中选择 18 名优秀队员参加竞赛.

- (1) 建立问题的层次结构图

层次结构图共分为三层:

最高层为目标(O)层:选择优秀队员;中间层为准则(C)层:队员的各项条件;最低层为方案(P)层:被选择的 20 名队员.

- (2) 确定准则(C)层对目标(O)层的权重

由问题中所给条件可知,选择队员主要依据的七项条件是依次排列的,Ⅰ、Ⅱ、Ⅲ、Ⅳ、Ⅴ、Ⅵ、Ⅶ七项条件对目标决策的影响程度也是依次排列的,且任意两项的影响程度之差可以认为基本相等. 因此,确定比较矩阵为

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 1/2 & 1 & 2 & 3 & 4 & 5 & 6 \\ 1/3 & 1/2 & 1 & 2 & 3 & 4 & 5 \\ 1/4 & 1/3 & 1/2 & 1 & 2 & 3 & 4 \\ 1/5 & 1/4 & 1/3 & 1/2 & 1 & 2 & 3 \\ 1/6 & 1/5 & 1/4 & 1/3 & 1/2 & 1 & 2 \\ 1/7 & 1/6 & 1/5 & 1/4 & 1/3 & 1/2 & 1 \end{bmatrix}$$

是 7 阶正互反矩阵.

求出矩阵 A 的最大特征值为 $\lambda_{\max} \approx 7.195\ 53$, 相应的特征向量并作归一化为

$$W_0 = (0.354\ 284, 0.239\ 927, 0.158\ 655, 0.103\ 624, 0.067\ 564\ 6, \\ 0.044\ 769\ 3, 0.031\ 174\ 6)^T,$$

即为准则层对目标层的权重.

随机一致性指标为 $RI = 1.36$, 一致性指标 $CI^{(1)} = \frac{\lambda_{\max} - 7}{7 - 1} \approx 0.032\ 588\ 3$, 于是
一致性比例指标为 $CR^{(1)} = \frac{CI^{(1)}}{RI} \approx 0.023\ 962 < 0.1$, 因此, 比较矩阵 A 是合理的.

(3) 确定方案(P)层对准则(C)层的权重

根据问题所给的条件和模型的假设(1)可知, 队员的各项条件充分地反映了每个队员的能力和水平, 因此, 可以利用每个队员(方案)的各项条件的比构造相应的比较矩阵.

设

$$W_k = (w_1^{(k)}, w_2^{(k)}, \dots, w_{20}^{(k)})^T$$

为准则 C_k (k 项条件) 的相关数据(表 13-1), 记 $a_{ij}^{(k)} = \frac{w_i^{(k)}}{w_j^{(k)}} (i, j = 1, 2, \dots, 20)$,
则 $C_k - P$ 的比较矩阵为 $A_k = (a_{ij}^{(k)})_{20 \times 20}$, 且 A_k 均为一致阵 ($k = 1, 2, \dots, 7$).

由于 A_k ($k = 1, 2, \dots, 7$) 的非零特征值为 $\lambda = 20$, 相应的特征向量取第一列向量, 即

$$(a_{11}^{(k)}, a_{21}^{(k)}, \dots, a_{20,1}^{(k)})^T = \left(\frac{w_1^{(k)}}{w_1^{(k)}}, \frac{w_2^{(k)}}{w_1^{(k)}}, \dots, \frac{w_{20}^{(k)}}{w_1^{(k)}} \right)^T \\ = (w_1^{(k)}, w_2^{(k)}, \dots, w_{20}^{(k)})^T \frac{1}{w_1^{(k)}} = \frac{W_k}{w_1^{(k)}}.$$

此与向量 W_k 仅差一个比例常数 $\frac{1}{w_1^{(k)}}$, 显然 W_k 也是 A_k 的特征向量.

将 A_k 的特征向量 W_k 归一化分别可以得到方案 P 层对准则 C_k 的权重, 计算结果如表 13-2.

表 13-2 P 层对 C 层的权重

| C_k | W_1 | W_2 | W_3 | W_4 | W_5 | W_6 | W_7 |
|-------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| P_A | 0.049 826 2 | 0.052 234 5 | 0.047 316 8 | 0.050 031 3 | 0.045 168 7 | 0.050 802 1 | 0.047 244 1 |
| P_B | 0.047 508 7 | 0.051 073 7 | 0.046 739 8 | 0.040 650 4 | 0.044 025 2 | 0.048 663 1 | 0.015 748 |
| P_C | 0.046 349 9 | 0.049 912 9 | 0.049 047 9 | 0.053 158 2 | 0.052 601 5 | 0.051 336 9 | 0.062 992 1 |
| P_D | 0.049 826 2 | 0.051 654 1 | 0.047 893 8 | 0.060 037 5 | 0.055 460 3 | 0.051 871 7 | 0.062 992 1 |
| P_E | 0.050 984 9 | 0.048 752 2 | 0.049 047 9 | 0.048 155 1 | 0.049 171 | 0.049 197 9 | 0.070 866 1 |
| P_F | 0.053 302 4 | 0.053 395 2 | 0.047 316 8 | 0.049 405 9 | 0.051 458 | 0.048 128 3 | 0.047 244 1 |
| P_G | 0.053 302 4 | 0.055 716 8 | 0.051 933 1 | 0.045 028 1 | 0.052 029 7 | 0.049 197 9 | 0.070 866 1 |
| P_H | 0.040 556 2 | 0.046 430 6 | 0.056 549 3 | 0.038 774 2 | 0.049 742 7 | 0.051 871 7 | 0.047 244 1 |
| P_I | 0.044 611 8 | 0.047 591 4 | 0.048 470 9 | 0.040 650 4 | 0.054 888 5 | 0.049 732 6 | 0.039 370 1 |
| P_J | 0.048 088 1 | 0.047 011 | 0.049 624 9 | 0.043 152 | 0.048 599 2 | 0.050 267 4 | 0.031 496 1 |
| P_K | 0.052 143 7 | 0.047 591 4 | 0.046 162 7 | 0.048 780 5 | 0.051 458 | 0.050 802 1 | 0.039 370 1 |
| P_L | 0.055 619 9 | 0.052 814 9 | 0.046 739 8 | 0.061 913 7 | 0.049 742 7 | 0.051 871 7 | 0.047 244 1 |
| P_M | 0.055 040 6 | 0.055 716 8 | 0.047 893 8 | 0.050 656 7 | 0.051 458 | 0.049 732 6 | 0.055 118 1 |
| P_N | 0.049 826 2 | 0.048 171 8 | 0.047 316 8 | 0.050 656 7 | 0.051 458 | 0.048 128 3 | 0.039 370 1 |
| P_O | 0.052 723 1 | 0.050 493 3 | 0.050 779 | 0.052 532 8 | 0.050 314 5 | 0.050 267 4 | 0.039 370 1 |
| P_P | 0.053 881 8 | 0.048 752 2 | 0.049 624 9 | 0.055 034 4 | 0.049 171 | 0.050 802 1 | 0.047 244 1 |
| P_Q | 0.048 667 4 | 0.046 430 6 | 0.054 241 2 | 0.057 536 | 0.048 027 4 | 0.048 663 1 | 0.055 118 1 |
| P_R | 0.050 405 6 | 0.048 171 8 | 0.053 087 1 | 0.056 910 6 | 0.049 742 7 | 0.049 197 9 | 0.062 992 1 |
| P_S | 0.045 191 2 | 0.047 011 | 0.055 395 3 | 0.047 529 7 | 0.051 458 | 0.051 336 9 | 0.070 866 1 |
| P_T | 0.052 143 7 | 0.051 073 7 | 0.054 818 2 | 0.049 405 9 | 0.044 025 2 | 0.048 128 3 | 0.047 244 1 |

(4) 确定方案 P 层对目标 O 层的组合权重

由准则 C 层对目标 O 层的权重 W_0 和方案 P 层对准则 C 层的权重, 可得方案 P 层对目标 O 层的组合权重为

$$W = (w_A, w_B, \dots, w_T)^T = [W_1, W_2, \dots, W_7] \cdot W_0.$$

经计算得

$$\begin{aligned}
 W &= (w_A, w_B, \dots, w_T)^T \\
 &= (0.049\ 675\ 6, 0.046\ 357\ 5, 0.049\ 602\ 8, 0.051\ 899\ 0, 0.050\ 265\ 9, \\
 &\quad 0.051\ 426\ 1, 0.053\ 084\ 8, 0.045\ 654\ 1, 0.046\ 288\ 6, 0.047\ 176\ 8, \\
 &\quad 0.049\ 249\ 4, 0.053\ 364\ 2, 0.053\ 137\ 3, 0.048\ 825\ 4, 0.051\ 170\ 9,
 \end{aligned}$$

$0.051\ 432\ 0, 0.050\ 091\ 7, 0.051\ 262\ 6, 0.048\ 988\ 0, 0.051\ 146\ 5)^T$, 其中向量 W 的 20 个分量分别为方案 P 层 (20 名队员) 相对目标层的权重.

由于方案层对准则层的比较矩阵 A_k 均为一致阵, 因此相应的

$$CI^{(2)} = (CI_1^{(2)}, CI_2^{(2)}, \dots, CI_7^{(2)}) \cdot W_0 = 0.$$

于是一致性比例指标 $CR^{(2)} = \frac{CI^{(2)}}{RI} = 0$, 因此组合一致性比例指标为

$$CR = CR^{(1)} + CR^{(2)} = 0.023\ 962 < 0.1$$

即组合权重可以作为目标决策的依据.

(5) 选择 18 名优秀队员

将如上方案 P 层对目标 O 层的权重作为每个队员的技术水平指标, 根据大小排序, 结果如表 13-3.

表 13-3 18 名队员的排序结果

| 序号 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| 队员 | L | W | G | D | F | P | R |
| 水平指标 | 0.053 364 2 | 0.053 137 3 | 0.053 084 8 | 0.051 899 0 | 0.051 426 1 | 0.051 432 0 | 0.051 262 6 |
| 序号 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 队员 | O | T | E | Q | A | C | K |
| 水平指标 | 0.051 170 9 | 0.051 146 5 | 0.050 265 9 | 0.050 091 7 | 0.049 675 6 | 0.049 602 8 | 0.049 249 4 |
| 序号 | 15 | 16 | 17 | 18 | 19 | 20 | |
| 队员 | S | N | J | B | I | H | |
| 水平指标 | 0.048 988 | 0.048 825 4 | 0.047 176 8 | 0.046 357 5 | 0.046 288 6 | 0.045 654 1 | |

由排序结果, 淘汰 H 和 I 两名队员, 其余的 18 名为入选参赛的优秀队员.

问题(2): 确定一个最佳的组队使竞赛技术水平最高.

由假设知, 每个队中的三名队员之间具有互补性, 即三人中各单项水平指标的最高者为该队的单项水平指标, 最佳组队主要体现全队在各单项水平指标最高, 不应有某项水平指标比其他的队低. 由问题(1)中的准则 C 层对目标 O 层的权重 W_0 可知, 七项准则是按顺序依次排列的, 对目标决策的影响是不同的, 而且前四项对目标决策起着决定性作用, 即水平指标主要体现在前四项上.

(1) 最佳组队原则

设 $m_i(x)$ 表示队员 x 的第 i 项水平指标, $M_i(x, y, z)$ 表示由队员 x, y, z 组队 (x, y, z) 的第 i 项水平指标, 则

$$M_i(x, y, z) = \max\{m_i(x), m_i(y), m_i(z)\} \quad (i = 1, 2, \dots, 7).$$

令 $M = (M_1(x, y, z), M_2(x, y, z), \dots, M_7(x, y, z))$, 于是用 $v(x, y, z) = M \cdot W_0$ 表示 (x, y, z) 队的整体技术水平指标. 最佳组队是从 18 名队员中选出 x, y, z 使 $v(x, y, z)$ 最高.

(2) 组队方案

根据组队原则, 最佳组队中的队员一定是前四项水平指标的最高者. 显然, 由表 13-2 可得 $m_1(L) = 0.055\ 619\ 9$ 为最高, 于是 $M_1 = m_1(L) = 0.055\ 619\ 9$, 则队员 L 是首先入选的队员. 其次 $m_2(G) = m_2(M) = 0.055\ 716\ 8$, 而 $m_3(G) > m_3(M)$, 故 $M_2 = m_2(G) = 0.055\ 716\ 8$, 则队员 G 是第二个入选的队员.

另外, $m_3(S) = 0.055\ 395\ 3$, 于是 $M_3 = m_3(S) = 0.055\ 395\ 3$, 而且 $M_4 = m_4(L) = 0.061\ 913\ 7$, 则队员 S 应是第三个入选的队员, 并且注意到 $M_5 = m_5(G) = 0.052\ 029\ 7$, $M_6 = m_6(L) = 0.051\ 871\ 7$, $M_7 = m_7(G) = 0.070\ 866\ 1$ 也都是相对的较高者, 即

$$M = (0.055\ 619\ 9, 0.055\ 716\ 8, 0.055\ 395\ 3, 0.061\ 913\ 7, \\ 0.052\ 029\ 7, 0.051\ 871\ 7, 0.070\ 866\ 1).$$

因此, 由队员 L, G, S 组成 (L, G, S) 队的技术水平指标为 $v(L, G, S) = M \cdot W_0 = 0.563\ 246$ 是最高的, 所以, 最佳组队为 (L, G, S) .

问题(3): 给出由 18 名队员组成 6 个队的组队方案, 使整体竞赛技术水平最高; 并给出每个队的竞赛技术水平.

根据每个队员的各项水平指标(表 13-2)和上面的假设及组队原则, 假设该问题是在保证最佳组队 (L, G, S) 不变的前提下, 考虑另外 15 名队员组成 5 个队使 6 个队整体技术水平指标最高.

(1) 问题的进一步分析

首先, 注意下面的基本事实: 最佳组队决策方案中的每个队对目标 O 层的权重一定不小于全体(18 名)队员对目标层权重的几何平均值, 否则其组队方案就不可能是最佳的; 三名队员的技术水平指标可以互补, 技术水平最高者为该队的水平指标.

由表 13-2, 按 $W_k = \left[\prod_{i=1}^{18} w_k^{(i)} \right]^{\frac{1}{18}}$ ($k = 1, 2, \dots, 7$) 计算可求出几何平均值, 且

$$W = (W_1, W_2, \dots, W_7) \cdot W_0 \\ = (0.051\ 012, 0.050\ 523, 0.049\ 902, 0.051\ 121, 0.049\ 919, \\ 0.050\ 160, 0.048\ 392) \cdot W_0.$$

任取三名队员 x, y, z 组合为一个队 (x, y, z) , 将对决策目标的权重定为该队的技术水平指标, 即 $v(x, y, z) = M \cdot W_0$, 此处 $M = (m_1, m_2, \dots, m_7)$, 其中七个分

量分别为 (x, y, z) 对准则 C 层的权重. 如果 $v(x, y, z) > W$, 则对应的 (x, y, z) 就可能是一个组队. 最后我们在所有队员的组合中确定出六个组队方案, 使每组技术水平指标之和为最高.

(2) 建立组队模型

用动态规划的方法, 分决策过程为 5 个阶段, 分步给出 5 个队的组队方案. 在除了队员 L, G, S 外的 15 名队员中组成 5 个队, 每一阶段确定一个队.

决策变量: $X_k = (x, y, z)_k (k = 1, 2, 3, 4, 5)$, 即任取三名队员 (x, y, z) 所组成的一个组队方案.

状态变量: $S_k (k = 1, 2, 3, 4, 5)$, 即从第 $k (1 \leq k \leq 5)$ 个到第 5 个组队的组队方案所包含的队员, 其中 $S_1 = \{A, B, C, \dots, T\}$.

状态转移方程: $S_{k+1} = S_k - X_k (k = 1, 2, 3, 4)$.

允许决策集合: $D_k = \{(x, y, z); x, y, z \in S_k, v_k(x, y, z) \geq W\} (k = 1, 2, 3, 4, 5)$.

指标函数: $v_k(S_k, X_k)$ 表示决策 X_k (一个组队) 关于状态 S_k 的技术水平指标, 即 $v_k(S_k, X_k) = M \cdot W_0$.

最优值函数: $f_k(S_k)$ 表示在状态 S_k 下确定的 $k (1 \leq k \leq 5)$ 个组队的技术水平指标之和的最大值. 则有逆序解法的基本方程:

$$\begin{cases} f_k(S_k) = \max_{x_k \in D_k} \{v_k(S_k, X_k) + f_{k+1}(S_{k+1})\}, & k = 5, 4, 3, 2, 1, \\ f_6(S_6) = v_6 = 0.056\ 324\ 6((L, G, S) \text{ 队的技术水平指标}), \end{cases}$$

其中 $S_{k+1} = S_k - X_k (k = 1, 2, 3, 4)$.

编程求解可以得到最优的组队方案 (如表 13-4), 其最优值为 $f_1(S_1) = 0.323\ 164$.

表 13-4 最优的组队方案

| 组队 X_k | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 |
|-------------------|-----------|-----------|-----------|-----------|-----------|-------------|
| 队员 (x, y, z) | (C, J, K) | (A, N, O) | (B, P, R) | (D, F, T) | (E, M, Q) | (G, L, S) |
| 水平 $v_k(x, y, z)$ | 0.051 647 | 0.051 935 | 0.053 262 | 0.054 647 | 0.055 348 | 0.056 324 6 |

13.5 参考案例与参考文献

1. 参考案例

- (1) 生产计划问题——文献[1]:375—382
- (2) 生产与库存问题——文献[2]:128—132
- (3) 工件加工问题——文献[3]:50

- (4) 基金分配问题——文献[4]:109—112
- (5) 不确定性采购问题——文献[5]:172—175

2. 参考文献

- [1] 姜启源. 数学模型. 第三版. 北京:高等教育出版社,2003
- [2] 白其峥. 数学建模案例分析. 北京:海洋出版社,2000
- [3] 袁震东. 数学建模方法. 上海:华东师范大学出版社,2003
- [4] 沈继红等. 数学建模. 哈尔滨:哈尔滨工程大学出版社,1996
- [5] 《数学模型》编写组. 数学模型. 广州:华南理工大学出版社,2003
- [6] 朱德通. 最优化模型与实验. 上海:同济大学出版社,2003
- [7] 何坚勇. 运筹学基础. 北京:清华大学出版社,1999
- [8] 刁在筠等. 运筹学. 北京:高等教育出版社,2001
- [9] 《运筹学》教材编写组. 运筹学. 修订版. 北京:清华大学出版社,1990
- [10] 韩中庚. 最佳组队方案及模型. 数学的实践与认识,1997,27(2):

第十四章 排队论方法

排队是日常生活中常见的一种现象,其共同的特点是:在一个排队服务系统中包含有一个或多个“服务设施”,有许多需要进入服务系统的“被服务者”,或“顾客”,当被服务者进入系统后不能立即得到服务,也就出现了排队现象.一个服务系统总是由“服务设施”与“被服务者”构成.如:医院与病人、商店与顾客、机场与飞机、火车站与火车、水库与水、网络与用户等等.由于“被服务者”到达服务系统的时间是不确定的,即是随机的,所以排队论又称为“随机服务系统理论”,因此,排队论在实际中有广泛的应用.

排队论要研究的内容有三部分:

- (1) 性态问题:即研究排队系统的概率分布规律,主要是研究队长分布、等待时间分布和忙期分布等;
- (2) 最优化问题:分为静态最优化和动态最优化,即为系统的最优设计和系统的最优运营问题;
- (3) 排队系统的统计推断:判断一个给定的排队系统符合于哪种模型,以便于根据排队理论进行分析研究.

14.1 排队论的基本概念

14.1.1 排队过程的一般模型

设要求服务的顾客从顾客总体进入排队系统(输入),到达服务机构前排队等候服务,服务完后立即离开(输出).

排队系统的主要有输入过程、排队规则和服务机构三个部分组成:

1. 输入过程:顾客到达排队系统的过程,具有如下的特征:

- (1) 顾客总体(称为顾客源)的组成可能是有限的,也可能是无限的;
- (2) 顾客到来的方式可能是一个一个的,也可能是成批的;
- (3) 顾客相继到达的间隔时间可以是确定型的,也可以是随机型的;
- (4) 顾客的到达是相互独立的;
- (5) 输入过程是平稳的,或称为对时间是齐次的,即相继到达的时间间隔分

布与时间无关.

2. 排队规则: 顾客到达后的排队方式、形状和队列数目,其特征有三条:

(1) 顾客到达后的排队方式可以是“即时制”,也可以是“等待制”,对于等待制的服务次序有:先到先服务、后到先服务、随机服务、有优先权的服务等;

(2) 排队可以有形的,也可以是无形的,有的排队容量是有限的,有的是无限的;

(3) 排队数目可以是单列,也可以是多列,有的可相互转移,有的不可相互转移.

3. 服务机构: 对顾客提供服务的设施或对象,从机构的形式和工作情况来分有以下特征:

(1) 服务机构可以没有服务员(或服务台),也可以有一个或多个服务台;

(2) 对于多个服务台的情况,可以是并列,可以是串列,也可以是混合排列;

(3) 服务方式可以是一个一个的进行,也可以是成批成批的进行;

(4) 服务时间可以是确定型的,也可以是随机型的,对于随机型的需要知道它的概率分布;

(5) 服务时间的分布对时间是平稳的,即分布均值、方差等都与时间无关.

14.1.2 排队模型的标准形式

排队模型的标准形式为 $X/Y/Z/A/B/C$, 其中 X 表示相继到达间隔时间的分布, Y 表示服务时间的分布, X 和 Y 的取值有下列几种情况:

M (Markov) 表示负指数分布;

D (Deterministic) 表示确定型的分布;

E_k (Erlang) 表示 k 阶爱尔朗分布;

GI (General Independent) 表示一般相互独立的时间间隔分布;

G (General) 表示一般服务时间的分布.

Z 表示服务台的个数, A 表示系统的容量限制, B 表示顾客源数目, C 表示服务规则,可分为先到先服务 (FCFS)、后到先服务 (LCFS)、随机服务、有优先权的服务等,通常只考虑 FCFS 的情况,此时可省略此项,例如: $M/M/C/N/m$, 即顾客相继到达间隔时间为负指数分布,服务时间为负指数分布, C 个服务台,系统容量为 N ,顾客源数为 m ,先到先服务.

14.1.3 排队系统的运行指标

排队论主要是研究排队系统运行的效率,估计服务质量,确定系统参数的最优值,以决定系统结构是否合理、研究设计改进措施. 因此,研究排队问题,首先要确定用以判断系统运行优劣的基本量化指标,然后求出这些指标的概率分布

和数学特征,要研究的系统运行指标主要有

- (1) 队长:指在系统中的顾客数,期望值记作 L_s ;
 - (2) 排队长(队列长):指在系统中排队等待服务的顾客数,其期望值记作 L_q ,即 $L_s = L_q + L_n$,其中 L_n 为正在接受服务的顾客数;
 - (3) 逗留时间:指一个顾客在系统中的停留时间,其期望值记作 W_s ;
 - (4) 等待时间:指一个顾客在系统中排队等待的时间,其期望值记作 W_q ,即 $W_s = W_q + \tau$,其中 τ 为服务时间;
 - (5) 忙期:服务机构连续工作的时间长度,记作 T_b ;
 - (6) 损失率:由于系统的条件限制,使顾客被拒绝服务而使服务部门受到损失的概率,用 $P_{\text{损}}$ 表示;
 - (7) 服务强度:
- 绝对通过能力 A ,表示单位时间内被服务完顾客的均值,或称为平均服务率;
- 相对通过能力 Q ,表示单位时间内被服务完的顾客数与请求服务的顾客数之比。

14.1.4 系统状态的概率

系统状态是求运行指标的基础,所谓系统的状态是指系统中顾客的数量.如果系统中有 n 个顾客,则说系统的状态为 n ,即可能的取值为

- (1) 当队长无限制时,则 $n = 0, 1, 2, \dots$;
- (2) 当队长为有限制,且最大值为 N 时,则 $n = 0, 1, 2, \dots, N$;
- (3) 当服务台个数为 c ,且服务为即时制时,则 $n = 0, 1, 2, \dots, c$.

一般说来,状态的取值与时间 t 有关,因此,在时刻 t 系统状态取值为 n 的概率记为 $P_n(t)$,如果 $\lim_{t \rightarrow \infty} P_n(t) = P_n$,则称为稳态(或统计平衡状态)解.实际中多数问题都有是属于稳态的情况,并不是真正的 $t \rightarrow \infty$,即过某一段时间以后就有 $P_n(t) \rightarrow P_n$.

14.2 到达时间的间隔分布和服务时间的分布

实际中,到达时间的间隔分布和服务时间的分布一般服从于以下三种分布:泊松(Poisson)分布、负指数分布和爱尔朗分布.

14.2.1 泊松分布

设 $N(t)$ 表示在时间段 $[0, t)$ 内到达的顾客数, $P_n(t_1, t_2)$ 表示在时间段 $[t_1, t_2)$ ($t_2 > t_1$) 内有 n ($n \geq 0$) 个顾客到达的概率,即 $P_n(t_1, t_2) = P\{N(t_2) -$

$N(t_1) = n$ }, 当 $P_n(t_1, t_2)$ 满足如下三个条件时, 则称顾客的到达形成泊松流:

(1) 无后效性: 在不相交的时间区间内顾客到达数是相互独立的, 即在时间区间 $[t, t + \Delta t]$ 内到达 k 个顾客的概率与时刻 t 以前到达多少顾客无关.

(2) 平稳性: 对于充分小的 Δt , 在时间间隔 $[t, t + \Delta t]$ 内有 1 个顾客到达的概率只与时间段的长度 Δt 有关, 而与起始时刻 t 无关, 且 $P_1(t, t + \Delta t) = \lambda \Delta t + o(\Delta t)$, 其中 $\lambda > 0$ 称为概率强度 (或平稳流强度), 即表示单位时间内有一个顾客到达的概率.

(3) 普通性: 对于充分小的 Δt , 在时间间隔 $[t, t + \Delta t]$ 内有 2 个或 2 个以上顾客的概率极小, 可以忽略不计, 即 $\sum_{n=2}^{\infty} P_n(t, t + \Delta t) = o(\Delta t)$.

下面研究系统状态为 n 的概率分布.

如果取时间段的初始时间为 $t = 0$, 则可记 $P_n(0, t) = P_n(t)$, 在 $[t, t + \Delta t]$ 内, 由于

$$\begin{aligned} \sum_{n=0}^{\infty} P_n(t, t + \Delta t) &= P_0(t, t + \Delta t) + P_1(t, t + \Delta t) \\ &\quad + \sum_{n=2}^{\infty} P_n(t, t + \Delta t) = 1, \end{aligned}$$

故在 $[t, t + \Delta t]$ 内没有顾客到达的概率为

$$\begin{aligned} P_0(t, t + \Delta t) &= 1 - P_1(t, t + \Delta t) - \sum_{n=2}^{\infty} P_n(t, t + \Delta t) \\ &= 1 - \lambda \Delta t + o(\Delta t). \end{aligned}$$

将 $[0, t + \Delta t]$ 分为 $[0, t]$ 和 $[t, t + \Delta t]$, 则在时间段 $[0, t + \Delta t]$ 内到达 n 个顾客的概率应为

$$\begin{aligned} P_n(t + \Delta t) &= P\{N(t + \Delta t) - N(0) = n\} \\ &= \sum_{k=0}^n P\{N(t + \Delta t) - N(t) = k\} \cdot P\{N(t) - N(0) = n - k\} \\ &= \sum_{k=0}^n P_{n-k}(t) \cdot P_k(t, t + \Delta t) \\ &= P_n(t)(1 - \lambda \Delta t) + P_{n-1}(t)\lambda \Delta t + o(\Delta t), \end{aligned}$$

即

$$\frac{P_n(t + \Delta t) - P_n(t)}{\Delta t} = -\lambda P_n(t) + \lambda P_{n-1}(t) + \frac{o(\Delta t)}{\Delta t}.$$

令 $\Delta t \rightarrow 0$, 则

$$\begin{cases} \frac{dP_n(t)}{dt} = -\lambda P_n(t) + \lambda P_{n-1}(t), \\ P_n(0) = 0 \quad (n \geq 1). \end{cases} \quad (14.1)$$

特别地,当 $n=0$ 时有

$$\begin{cases} \frac{dP_0(t)}{dt} = -\lambda P_0(t), \\ P_0(0) = 1. \end{cases} \quad (14.2)$$

由(14.2)式解得 $P_0(t) = e^{-\lambda t}$,代入(14.1)式解得 $P_1(t) = \lambda t e^{-\lambda t}$, $P_2(t) = \frac{(\lambda t)^2}{2!} e^{-\lambda t}$.

一般地有 $P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (n = 0, 1, 2, \dots, t > 0)$

表示在长为 t 的时间段内到达 n 个顾客的概率,即为泊松分布,数学期望和方差分别为 $E[N(t)] = \lambda t$, $D[N(t)] = \lambda t$.

14.2.2 负指数分布

当顾客流为泊松流时,用 T 表示两个相继到达的时间间隔,是一个随机变量,其分布函数 $F_T(t) = P\{T \leq t\} = 1 - P\{T > t\} = 1 - P_0(t)$. 由上可知 $P_0(t) = e^{-\lambda t}$,于是 $F_T(t) = 1 - e^{-\lambda t}$, $t \geq 0$, 分布密度为 $f_T(t) = \lambda e^{-\lambda t}$, $t \geq 0$. 这里 λ 表示单位时间内平均到达的顾客数,则 $E(T) = \frac{1}{\lambda}$ 表示相继顾客到达平均间隔时间.

类似地,设系统对一个顾客的服务时间为 v (即在忙期内相继离开系统的两个顾客的间隔时间)服从于负指数分布,分布函数为 $F_v(t) = 1 - e^{-\mu t}$, $t \geq 0$, 分布密度为 $f_v(t) = \mu e^{-\mu t}$, $t \geq 0$, 其中 μ 表示平均服务率 (即单位时间内被服务完的顾客数),且期望值为 $E(v) = \frac{1}{\mu}$, 表示平均一个顾客的服务时间.

因此, T 服从于负指数分布,即与概率强度为 λ 的泊松流等价. 并且注意到,由条件概率可知: $P\{T > t + s | T > s\} = P\{T > t\}$, 即说明后一个顾客的到来所需时间与前一个顾客到来所需时间 s 无关,故 T 具有无记忆性. 于是,在排队模型的记号中都用 M 表示,且 $E(T) = \frac{1}{\lambda}$, $D(T) = \frac{1}{\lambda^2}$.

14.2.3 爱尔朗分布

设有如下的顾客流,记 k 个顾客到达系统的时间间隔序列为 v_1, v_2, \dots, v_k (为相

互独立的随机变量), 同服从于参数为 $k\mu$ 的负指数分布, 则称随机变量 $T = \sum_{i=1}^k v_i$

服从于 k 阶爱尔朗分布, 分布密度为 $f_k(t) = \frac{\mu k (\mu k t)^{k-1}}{(k-1)!} e^{-\mu k t} \quad (t > 0, \mu > 0)$.

$$\begin{aligned} \text{注: } E[v_i] &= \frac{1}{k\mu} \quad (i = 1, 2, \dots, k), \text{ 则 } E(T) = \sum_{i=1}^k E[v_i] = \frac{1}{\mu}, \text{ 且 } D(T) \\ &= \frac{1}{k\mu^2}. \end{aligned}$$

当 $k=1$ 时, 即为负指数分布, 因此爱尔朗分布比负指数分布更广泛.

类似地, 设系统中有串列的 k 个服务台, 每个服务台对顾客的服务时间是相互独立的, 且服从于 (参数为 $k\mu$) 负指数分布, 则一个顾客在接受完 k 个服务台的服务所需的总时间 T 服从于 k 阶爱尔朗分布.

14.3 单服务台的排队模型

设系统的输入过程服从于泊松流, 服务时间服从于负指数分布, 单服务台的排队系统有以下三种情形:

- (1) 标准型: $M/M/1(M/M/1/\infty/\infty)$;
- (2) 系统容量有限制: $M/M/1/N/\infty$;
- (3) 顾客源为有限的: $M/M/1/\infty/m$.

14.3.1 标准型: $M/M/1$

排队模型 $M/M/1$ 表示顾客源为无限的, 顾客的到达相互独立, 到达规律服从参数 λ 的泊松分布; 单服务台、队长无限、先到先服务; 各顾客的服务时间相互独立, 且同服从于参数为 μ 的负指数分布.

1. 确定系统在任意时刻 t 的状态为 n 的概率

因为已知顾客的到达规律服从参数为 λ 的泊松分布, 服务时间服从参数为 μ 的负指数分布, 于是在时间间隔 $[t, t + \Delta t)$ 内有

- (1) 有一个顾客到达的概率为 $\lambda \Delta t + o(\Delta t)$;
- (2) 没有一个顾客到达的概率为 $1 - \lambda \Delta t + o(\Delta t)$;
- (3) 有一个顾客被服务完的概率为 $\mu \Delta t + o(\Delta t)$;
- (4) 没有一个顾客被服务完的概率为 $1 - \mu \Delta t + o(\Delta t)$;
- (5) 多于一个顾客到达或被服务完离去的概率为 $o(\Delta t)$.

现在考虑在 $t + \Delta t$ 时刻系统中有 n 个顾客 (即状态为 n) 的概率 $P_n(t + \Delta t)$, 可能的情况如表 14-1.

表 14-1 系统状态的变化规律

| 情况 | 时刻 t 的 顾客数 | 在区间 $(t, t + \Delta t)$ | | 在时刻 $t + \Delta t$ 的顾客数 | $P_n(t + \Delta t)$ |
|-----|-----------------|-------------------------|-----|----------------------------|--|
| | | 到 达 | 离 去 | | |
| (A) | n | × | × | n | $P_n(t)(1 - \lambda\Delta t)(1 - \mu\Delta t)$ |
| (B) | $n + 1$ | × | ✓ | n | $P_{n+1}(t)(1 - \lambda\Delta t)\mu\Delta t$ |
| (C) | $n - 1$ | ✓ | × | n | $P_{n-1}(t)(\lambda\Delta t)(1 - \mu\Delta t)$ |
| (D) | n | ✓ | ✓ | n | $P_n(t)(\lambda\Delta t)(\mu\Delta t)$ |

这是一个生灭过程,四种情况是相互独立的事件. 则有

$$P_n(t + \Delta t) = P_n(t)(1 - \lambda\Delta t - \mu\Delta t) + P_{n+1}(t)\mu\Delta t + P_{n-1}(t)\lambda\Delta t + o(\Delta t).$$

移项整理,两边同除以 Δt ,并令 $\Delta t \rightarrow 0$,则得

$$\frac{dP_n(t)}{dt} = \lambda P_{n-1}(t) + \mu P_{n+1}(t) - (\lambda + \mu)P_n(t), \quad n = 1, 2, \dots$$

$$\text{当 } n=0 \text{ 时,类似地可有 } \frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t).$$

于是,一般地有

$$\begin{cases} \frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t), \\ \frac{dP_n(t)}{dt} = \lambda P_{n-1}(t) + \mu P_{n+1}(t) - (\lambda + \mu)P_n(t), \quad n > 1. \end{cases}$$

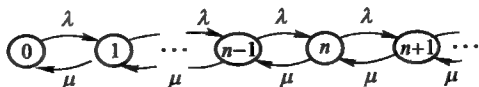
此方程为差分微分方程,其解是贝塞尔(Bessell)函数,不便于应用,为此,我们只研究它的稳态解. 假设当 $t \rightarrow \infty$ 时,极限存在,或 $P_n(t)$ 与 t 无关,记 $P_n(t)$ 为

P_n , 即有 $\frac{dP_n(t)}{dt} = 0$, 则

$$\begin{cases} -\lambda P_0 + \mu P_1 = 0, \\ \lambda P_{n-1} + \mu P_{n+1} - (\lambda + \mu)P_n = 0, \quad n > 1 \end{cases} \quad \text{或} \quad \begin{cases} \lambda P_0 = \mu P_1, \\ \lambda P_{n-1} + \mu P_{n+1} = (\lambda + \mu)P_n. \end{cases}$$

这是关于 P_n 的差分方程,也反映出了系统状态的转移关系,即每一状态都是平

衡的,如图 14-1. 求得 $P_1 = \left(\frac{\lambda}{\mu}\right)P_0$, 递推可得 $P_n = \left(\frac{\lambda}{\mu}\right)^n P_0 (n \geq 1)$.

图 14-1 $M/M/1$ 的状态转移关系图

令 $\rho = \frac{\lambda}{\mu} < 1$, 称为服务强度, 即为平均到达率与平均服务率之比, 由概率

的性质: $\sum_{n=0}^{\infty} P_n = 1$, 即 $P_0 \sum_{n=0}^{\infty} \rho^n = \frac{P_0}{1-\rho} = 1$, 于是

$$\begin{cases} P_0 = 1 - \rho, \\ P_n = (1 - \rho)\rho^n, n \geq 1, \end{cases}$$

就是所求的系统状态为 n 的概率.

2. 系统的运行指标

(1) 队长(平均顾客数): 因为系统的状态为 n , 即系统中有 n 个顾客, 由期望的定义得

$$L_s = \sum_{n=0}^{\infty} nP_n = \sum_{n=1}^{\infty} n(1-\rho)\rho^n = \frac{\rho}{1-\rho} = \frac{\lambda}{\mu-\lambda};$$

(2) 队列长(等待的平均顾客数): $L_q = \sum_{n=1}^{\infty} (n-1)P_n = L_s - \rho = \frac{\rho^2}{1-\rho} =$

$$\frac{\rho\lambda}{\mu-\lambda};$$

(3) 系统中顾客的逗留时间: 事实上, 系统中的一个顾客的逗留时间 w , 服从于参数为 $(\mu-\lambda)$ 的负指数分布, 分布函数和分布密度分别为

$$F(w) = 1 - e^{-(\mu-\lambda)w}, f(w) = (\mu-\lambda)e^{-(\mu-\lambda)w},$$

所以有 $w_s = E(w) = \frac{1}{\mu-\lambda}$;

(4) 系统中顾客的等待时间: 等待时间 = 逗留时间 - 被服务的时间, 即

$$W_q = W_s - \frac{1}{\mu} = \frac{\rho}{\mu-\lambda},$$

其中 $\frac{1}{\mu}$ 表示平均一个顾客的服务时间.

综上所述, 系统主要运行指标为

$$L_s = \frac{\lambda}{\mu-\lambda}, L_q = \frac{\rho\lambda}{\mu-\lambda}, w_s = \frac{1}{\mu-\lambda}, W_q = \frac{\rho}{\mu-\lambda}.$$

3. 运行指标之间的关系

$$L_s = \lambda w_s, L_q = \lambda w_q, w_s = w_q + \frac{1}{\mu}, L_s = L_q + \frac{\lambda}{\mu},$$

这些关系式称之为 Little 公式.

14.3.2 系统的容量有限制: $M/M/1/N/\infty$

1. 系统的状态概率

类似于标准型的假设, 不同的是队长有限制 ($n \leq N$), 即系统中最多允许有

N 个顾客在排队,多了将被拒绝进入系统. 假设 λ 为平均到达率, μ 为平均服务率, 则类似地可得状态转移关系图, 如图 14-2.

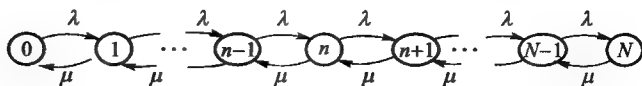


图 14-2 $M/M/1/N/\infty$ 的状态转移关系图

于是状态概率的(稳态)方程为

$$\begin{cases} \mu P_1 = \lambda P_0, \\ \mu P_{n+1} + \lambda P_{n-1} = (\lambda + \mu) P_n \quad (1 \leq n \leq N-1), \\ \mu P_N = \lambda P_{N-1}. \end{cases}$$

注意到 $\sum_{n=0}^N P_n = 1$, 由递推关系不难求得系统的状态概率为

$$\begin{cases} P_0 = \frac{1 - \rho}{1 - \rho^{N+1}}, \rho \neq 1, \\ P_n = \frac{1 - \rho}{1 - \rho^{N+1}} \rho^n, 1 \leq n \leq N. \end{cases}$$

注:

(1) 如果 $\rho = 1$ (即 $\lambda = \mu$), 由 $P_0 = P_1 = \cdots = P_N = \frac{1}{N+1}$, 即到达率和服务率相等, 在稳态情况下系统不会出现排队等待现象;

(2) 这里因为是有有限项的和, 所以不要求 $\rho < 1$, 但当 $\rho = \lambda/\mu > 1$, 即 $\lambda > \mu$ 时, 表示单位时间内的到达率比服务率大, 系统的损失率增加, 即被拒绝排队的数量增大.

2. 系统的运行指标

$$(1) \text{ 队长: } L_s = \sum_{n=0}^N n P_n = \frac{\rho}{1 - \rho} - \frac{(N+1)\rho^{N+1}}{1 - \rho^{N+1}}, \rho \neq 1;$$

$$(2) \text{ 队列长: } L_q = \sum_{n=1}^N (n-1) P_n = L_s - (1 - P_0);$$

(3) 顾客逗留时间: 首先注意到一个事实, 因为 w_s 与平均到达率 λ 有关, 而 λ 表示系统的容量有空时的平均到达率, 当系统满员 ($n = N$) 时, 则到达率为 0, 为此引入有效到达率 λ_e , 表示有空时平均到达率 λ 减去满员后拒绝顾客的平均数 λP_N , 即 $\lambda_e = \lambda(1 - P_N)$, 由于

$$\lambda_e = \lambda(1 - P_N) = \lambda \left(1 - \frac{1 - \rho}{1 - \rho^{N+1}} \rho^N \right) = \lambda \frac{1 - \rho^N}{1 - \rho^{N+1}}$$

和

$$\mu = \frac{\lambda P_0}{P_1} = \frac{1 - \rho^{N+1}}{(1 - \rho)\rho} \lambda P_0 = \frac{\lambda}{\rho},$$

所以有效服务强度为

$$\frac{\lambda_e}{\mu} = \lambda \left(1 - \frac{1 - \rho^N}{1 - \rho^{N+1}} \right) \frac{\rho}{\lambda} = \frac{\rho - \rho^{N+1}}{1 - \rho^{N+1}} = 1 - P_0,$$

于是顾客逗留时间为 $w_s = \frac{L_s}{\lambda_e} = \frac{L_s}{\lambda(1 - P_0)} = \frac{L_q}{\lambda(1 - P_N)} + \frac{1}{\mu}$;

$$(4) \text{ 顾客等待时间: } w_q = \frac{L_q}{\lambda_e} = \frac{L_q}{\lambda(1 - P_N)} = w_s - \frac{1}{\mu}.$$

14.3.3 顾客源为有限的: $M/M/1/\infty/m$

对该模型的顾客总体虽只有 m 个顾客,但每个顾客的到来并接受服务后,仍然回到顾客总体,即可以再次到来,所以对系统的容量是没有限制的,实际上系统中顾客数永远不会超过 m ,即与模型 $M/M/1/m/m$ 的意义相同。

与前面情况类似,假设每个顾客的到达率相同为 λ ,在系统外的平均顾客数为 $m - L_s$,故系统的有效到达率为 $\lambda_e = \lambda(m - L_s)$ 。考虑稳态的情况,则系统的状态转移关系如图 14-3。

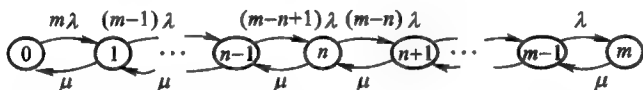


图 14-3 $M/M/1/\infty/m$ 的状态转移关系图

于是可得系统状态概率的平衡方程为

$$\begin{cases} \mu P_1 = m\lambda P_0, \\ \mu P_{n+1} + (m - n + 1)\lambda P_{n-1} = [(m - n)\lambda + \mu]P_n \quad (1 \leq n \leq m - 1), \\ \mu P_m = \lambda P_{m-1}. \end{cases}$$

注意到 $\sum_{n=0}^m P_n = 1$, 由递推关系不难求得系统状态的概率为

$$\begin{cases} P_0 = \frac{1}{\sum_{i=0}^m \frac{m!}{(m-i)!} \left(\frac{\lambda}{\mu}\right)^i}, \\ P_n = \frac{m!}{(m-n)!} \left(\frac{\lambda}{\mu}\right)^n P_0, \quad 1 \leq n \leq m. \end{cases}$$

系统的运行指标为

$$L_s = \sum_{n=0}^m nP_n = m - \frac{\mu}{\lambda}(1 - P_0), \quad w_s = \frac{m}{\mu(1 - P_0)} - \frac{1}{\lambda}, \quad w_q = w_s - \frac{1}{\mu},$$

$$L_q = \sum_{n=1}^m (n-1)P_n = m - \frac{(\lambda + \mu)(1 - P_0)}{\lambda} = L_s - (1 - P_0).$$

14.4 多服务台的排队模型

研究单队、并列的 c 个服务台的情形,主要有三种情形:

- (1) 标准型: $M/M/c(M/M/c/\infty/\infty)$;
- (2) 系统容量有限制: $M/M/c/N/\infty$;
- (3) 顾客源为有限的: $M/M/c/\infty/m$.

14.4.1 标准型: $M/M/c/\infty/\infty$

前提假设同 $M/M/1/\infty/\infty$, 另外, 假设顾客流为泊松流, 平均到达率为 λ , 各服务台的服务时间满足负指数分布, 而各服务台的工作是相互独立的(不搞协作), 单个服务台的平均服务率为 μ , 则整个服务机构的平均服务率为 $c\mu$ (当 $n \geq c$), 或 $n\mu$ (当 $n < c$), 令 $\rho = \frac{\lambda}{c\mu}$, 称为系统的服务强度(服务机构的平均利用率), 当 $\rho > 1$ 时, 系统就会出排队现象, 即有顾客在排队等待.

类似地, 可以得到系统状态概率的平衡方程为

$$\begin{cases} \mu P_1 = \lambda P_0, \\ (n+1)\mu P_{n+1} + \lambda P_{n-1} = (\lambda + n\mu)P_n \quad (1 \leq n \leq c), \\ c\mu P_{n+1} + \lambda P_{n-1} = (\lambda + c\mu)P_n \quad (n > c), \end{cases}$$

其中 $\sum_{n=0}^{\infty} P_n = 1$, 且 $\rho = \frac{\lambda}{c\mu} \leq 1$. 由递推关系可以求得系统状态概率

$$P_0 = \left[\sum_{k=0}^{c-1} \frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k + \frac{1}{c!} \frac{1}{1-\rho} \left(\frac{\lambda}{\mu} \right)^c \right]^{-1}, P_n = \begin{cases} \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n P_0, & \text{当 } n \leq c, \\ \frac{1}{c!} \frac{1}{c^{n-c}} \left(\frac{\lambda}{\mu} \right)^n P_0, & \text{当 } n > c. \end{cases}$$

系统的运行指标为

$$L_s = L_q + \frac{\lambda}{\mu},$$

$$L_q = \sum_{n=c+1}^{\infty} (n-c)P_n = \sum_{k=1}^{\infty} kP_{k+c} = \sum_{k=1}^{\infty} \frac{k}{c!c^k} (c\rho)^{k+c} P_0 = \frac{(c\rho)^c \rho}{c!(1-\rho)^2} P_0,$$

$$\text{其中 } c\rho = \frac{\lambda}{\mu}, k = n - c, w_s = \frac{L_s}{\lambda} = \frac{(c\rho)^c \rho}{c!(1-\rho)^2 \lambda} P_0 + \frac{1}{\mu}, w_q = \frac{L_q}{\lambda} = \frac{(c\rho)^c \rho}{c!(1-\rho)^2 \lambda} P_0.$$

思考题: 一个 $M/M/c/\infty/\infty$ 系统与 c 个 $M/M/1/\infty/\infty$ 系统比较哪一种效率高?

14.4.2 系统容量有限制: $M/M/c/N/\infty$

假设系统内有 c 个服务台, 顾客流为泊松流, 平均到达率为 λ . 各服务台的服务时间服从负指数分布, 而工作是相互独立的, 平均服务率为 μ . 系统的最大容量为 $N (N \geq c)$, 当系统客满 (即系统内有 N 个顾客) 时, 有 c 个接受服务, $N - c$ 个在排队, 再有顾客到来将被拒绝而离去, 系统将有损失率.

当系统的状态为 n 时, 每个服务台的服务率为 μ , 则系统的总服务率: 当 $0 < n < c$ 时为 $n\mu$; 当 $n \geq c$ 时为 $c\mu$, 令 $\rho = \frac{\lambda}{c\mu}$ 为系统的服务强度.

类似地, 可以得到系统的状态概率平衡方程

$$\begin{cases} \mu P_1 = \lambda P_0, \\ (n+1)\mu P_{n+1} + \lambda P_{n-1} = (\lambda + n\mu) P_n \quad (1 \leq n \leq c), \\ c\mu P_{n+1} + \lambda P_{n-1} = (\lambda + c\mu) P_n \quad (c \leq n < N), \\ \lambda P_{N-1} = c\mu P_N, \end{cases}$$

其中 $\sum_{n=0}^N P_n = 1$, 且 $\rho = \frac{\lambda}{c\mu} \leq 1$. 由递推关系可以求得系统状态概率为

$$P_0 = \begin{cases} \left[\sum_{k=0}^{c-1} \frac{1}{k!} (c\rho)^k + \frac{c^c}{c!} \frac{\rho(\rho^c - \rho^N)}{1 - \rho} \right]^{-1}, & \rho \neq 1, \\ \sum_{k=0}^{c-1} \frac{1}{k!} (c)^k + \frac{c^c}{c!} (N - c + 1), & \rho = 1, \end{cases}$$

$$P_n = \begin{cases} \frac{1}{n!} (c\rho)^n P_0, & \text{当 } 0 \leq n \leq c, \\ \frac{c^c}{c!} \rho^{n-c} P_0, & \text{当 } c < n \leq N. \end{cases}$$

系统的运行指标为

$$L_s = L_q + c\rho(1 - P_N), \quad w_s = w_q + \frac{1}{\mu}, \quad w_q = \frac{L_q}{\lambda(1 - P_N)},$$

$$L_q = \sum_{n=c+1}^N (n - c) P_n = \frac{(c\rho)^c \rho}{c! (1 - \rho)^2} P_0 [1 - \rho^{N-c} - (N - c)(1 - \rho)\rho^{N-c}],$$

系统满员的损失率为 $P_{\text{损}} = P_N = \frac{c^c}{c!} \rho^N P_0$.

特别地,当 $N=c$ 时,即 $M/M/C/C/\infty$,此时系统为即时制服务,不允许有顾客在系统内排队,亦即系统的状态概率为

$$P_0 = \left[\sum_{k=0}^{c-1} \frac{(c\rho)^k}{k!} \right]^{-1}, P_n = \frac{(c\rho)^n}{c!} P_0,$$

相应地有运行指标为

$$L_q = W_q = 0, W_s = \frac{1}{\mu}, L_s = \sum_{n=1}^c nP_n = c\rho(1 - P_c).$$

14.4.3 顾客源为有限的: $M/M/c/\infty/m$

假设同前面的问题相同,即 c, m, λ, μ, ρ 的意义相同,系统状态概率的平衡方程为

$$\begin{cases} \mu P_1 = m\lambda P_0, \\ (n+1)\mu P_{n+1} + (m-n+1)\lambda P_{n-1} = [(m-n)\lambda + n\mu]P_n \quad (1 \leq n \leq c), \\ c\mu P_{n+1} + (m-c+1)\lambda P_{n-1} = [(m-c)\lambda + c\mu]P_n \quad (c \leq n < m), \\ \lambda P_{m-1} = c\mu P_m. \end{cases}$$

由递推关系可以求得状态概率

$$P_0 = \frac{1}{m!} \left[\sum_{k=0}^c \frac{1}{k!(m-k)!} \left(\frac{c\rho}{m} \right)^k + \frac{c^c}{c!} \sum_{k=0}^c \frac{1}{(m-k)!} \left(\frac{\rho}{m} \right)^k \right]^{-1}, \rho = \frac{m\lambda}{c\mu},$$

$$P_n = \begin{cases} \frac{m!}{(m-n)!n!} \left(\frac{\lambda}{\mu} \right)^n P_0, & \text{当 } 0 \leq n \leq c, \\ \frac{m!}{(m-n)!c!c^{n-c}} \left(\frac{\lambda}{\mu} \right)^n P_0, & \text{当 } c < n \leq m. \end{cases}$$

系统的运行指标为

$$L_s = \sum_{n=1}^m nP_n, L_q = \sum_{n=c+1}^m (n-c)P_n, w_s = \frac{L_s}{\lambda_e}, w_q = \frac{L_q}{\lambda_e},$$

有效到达率为 $\lambda_e = \lambda(m - L_s)$, 且 $L_s = L_q + \frac{\lambda_e}{\mu} = L_q + \frac{\lambda}{\mu}(m - L_s)$.

类似地还有 $M/M/c/N/m, M/M/c/m/m, M/M/c/c/m$ 等情况,可作相应的讨论.

14.5 排队系统的最优化问题

14.5.1 一般排队系统的最优化问题

1. 最优化问题的分类

(1) 系统设计最优化,或称**静态最优化**,是指在服务系统设置以前根据一定的质量指标,找出参数的最优值,从而使系统设计最经济.例如:服务机构的规模大小、服务台的个数、系统容量大小等.

(2) 系统控制最优化,或称**动态最优化**,是指对已有的排队系统寻求使其某一目标函数达到最优的运营机制.

2. 费用模型

通常所说有费用是指服务机构的服务费用和顾客等待的费用,一般说来,提高服务机构的服务水平(即增加了服务机构的成本),自然会降低顾客的等待费用(损失),最优化的目标之一是使二者费用之和为最小,另一个目标是使服务机构的纯收入(利润)为最大.如图 14-4.

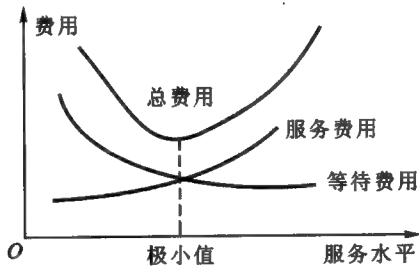


图 14-4 系统费用

14.5.2 模型 $M/M/1$ 中的最优服务率 μ

1. 标准型: $M/M/1$

设目标函数: $z = c_s \mu + c_w L_s$, 即为单位时间服务成本与顾客等待费用之和的期望值, 其中 c_s 表示当 $\mu = 1$ (单位时间内服务完一个顾客) 时服务机构的服务费用, c_w 为每个顾客在系统中停留单位时间的费用, 由 $L_s = \frac{\lambda}{\mu - \lambda}$, 则 $z = c_s \mu + c_w \frac{\lambda}{\mu - \lambda}$, 求其极小值. 即令 $\frac{dz}{d\mu} = 0$, 则 $c_s - \frac{c_w \lambda}{(\mu - \lambda)^2} = 0$, 解出最优解 $\mu^* = \lambda + \sqrt{\frac{c_w}{c_s} \lambda}$, 即为最优服务率.

2. 系统容量有限的: $M/M/1/N/\infty$

如果系统中已有 N 个顾客, 则后来的顾客将被拒绝, 于是可设 P_N 为被拒绝的概率, $1 - P_N$ 即为接受服务的概率. $\lambda(1 - P_N)$ 表示单位时间内实际进入服务机构的顾客数, 在稳定状态下, 即为单位时间内实际服务完成的顾客数.

设系统服务完 1 个顾客能收入 G 元, 于是单位时间收入的期望值为 $\lambda(1 - P_N)G$, 则系统的纯利润为

$$z = \lambda(1 - P_N)G - c_s\mu = \lambda G \frac{1 - \rho^N}{1 - \rho^{N+1}} - c_s\mu = \lambda G \frac{\mu^N - \lambda^N}{\mu^{N+1} - \lambda^{N+1}} - c_s\mu.$$

令 $\frac{dz}{d\mu} = 0$, 可解得

$$\rho^{N+1} \frac{N - (N+1)\rho + \rho^{N+1}}{(1 - \rho^{N+1})^2} = \frac{c_s}{G},$$

其中 $P_N = \frac{\rho^N - \rho^{N+1}}{1 - \rho^{N+1}}$, $\rho = \frac{\lambda}{\mu}$, 而 c_s, G, λ, N 均为已知的, 用数值方法求解出 μ^* 的数值解.

3. 顾客源为有限的: $M/M/1/\infty/m$

设顾客数为 m , 单个服务台、服务时间服从负指数分布, 当服务率为 $\mu = 1$ 时, 服务机构的成本费为 c_s , 单位时间内服务完一个顾客的收入为 G 元, 单位时间内服务完的顾客数为 $m - L_s$, 则单位时间内的纯利润为

$$z = (m - L_s)G - c_s\mu = \frac{mG}{\rho} \cdot \frac{E_{m-1}\left(\frac{m}{\rho}\right)}{E_m\left(\frac{m}{\rho}\right)} - c_s\mu,$$

其中 $E_m\left(\frac{m}{\rho}\right) = \sum_{k=1}^m \frac{\left(\frac{m}{\rho}\right)^k}{k!} e^{-\frac{m}{\rho}}$ 为泊松和, $\rho = \frac{m\lambda}{\mu}$, 令 $\frac{dz}{d\mu} = 0$, 则得

$$\frac{E_{m-1}\left(\frac{m}{\rho}\right)E_m\left(\frac{m}{\rho}\right) + \frac{m}{\rho}\left[E_m\left(\frac{m}{\rho}\right)E_{m-1}\left(\frac{m}{\rho}\right) - E_{m-1}^2\left(\frac{m}{\rho}\right)\right]}{E_m^2\left(\frac{m}{\rho}\right)} = \frac{c_s\lambda}{G}.$$

当给定 c_s, G, λ, m 后, 利用泊松分布表和数值方法计算求解得最优服务率 μ^* .

14.5.3 模型 $M/M/c$ 中的最优服务台数

仅对标准的模型进行讨论. 在稳态的假设下, 单位时间内每服务台的成本费

为 c_s , 每个顾客在系统中停留单位时间的费用为 c_w , 则单位时间内的费用(服务成本和等待的费用)的期望值: $z = c_s c + c_w L_s$, 其中 $L_s = L_s(c)$, 即与服务台的个数 c 有关, 因此总费用为 $z = z(c)$, 记 c 的最优值为 c^* , 则 $z(c^*)$ 是最小费用. 由于 c 只能取整数, 即 $z(c)$ 是离散函数, 所以只能用边际分析方法求解. 事实上, 根据 $z(c^*)$ 为最小值, 可有

$$\begin{cases} z(c^*) \leq z(c^* - 1), \\ z(c^*) \leq z(c^* + 1). \end{cases}$$

由 $z = c_s c + c_w L_s$, 则有

$$\begin{cases} c_s c^* + c_w L_s(c^*) \leq c_s(c^* - 1) + c_w L_s(c^* - 1), \\ c_s c^* + c_w L_s(c^*) \leq c_s(c^* + 1) + c_w L_s(c^* + 1), \end{cases}$$

化简整理得

$$L_s(c^*) - L_s(c^* + 1) \leq \frac{c_s}{c_w} \leq L(c^* - 1) - L_s(c^*).$$

14.6 校园网的设计和调节收费问题^[10]

14.6.1 问题的提出

随着计算机技术的飞速发展, 校园信息网已在全国高校中普及. 某高校拟建一校园信息网, 并与 internet 连接, 用户可以通过网络通信端口拨号上网, 为此, 需要根据用户的数量研究通信端口的设计规模. 通常的通信端口分为 16 口、32 口、64 口、128 口等, 实际中随着通信端口数量的增加, 其成本费将成倍增加. 如何根据实际情况在保证基本满足用户需求的前提下, 确定合适的通信端口数, 以减少费用开支和资源的浪费.

当网络建成后, 为了保证用户有效地使用信息网, 必须要通过适当的收取线路调节费来控制上网时间, 一般认为采用分段计时收费较为合理, 例如按上网时间长短分为“免费→半费→全费→2 倍→3 倍→4 倍……”等时段.

现在的问题是

(1) 假设有 m 个用户, 每个用户平均每天(按 16 h 计)上网 1.5 h, 试确定通信端口数 n 与 m 之比 n/m ;

(2) 假设 $m = 150$, 按所设定的通信端口数 n , 试讨论平均每天每个用户上网 1 h、1.5 h、2 h、3 h、4 h、5 h 的可能性, 出现因线路忙用户想上网而上不去所产生抱怨的可能性和通信端口的平均使用率;

(3) 为了控制上网时间, 学校要求适当收取线路调节费, 试给出一种合理的分段计时收取线路调节费的方案.

14.6.2 问题的分析与假设

根据问题中所给的信息,我们可以用排队理论来研究这一问题.假设校园信息网络和用户构成一个排队系统,网络的通信端口为服务台,个数为 n ,用户为顾客,顾客源数为 m ,平均忙期(即一天连续工作时间)为16 h.同时要注意到:实际中不限制用户上网的次数,虽然实际用户数为 m ,但我们可以认为顾客总体是无限的.

另一方面,在同一时间当 n 个通信端口全部被占用(即系统满员)时,再有用户拨号上网,系统将会拒绝,使用户无法上网.此时,这些用户会产生抱怨,只有在网上的用户下网后才能有新的用户上网,可以这样周而复始地进行下去.也就是说,只要时间允许,系统不限制上网的人数,但不允许用户在系统内排队等候,即系统服务是即时制的.为此,给出如下几点假设:

- (1) 每个用户的上网是随机、且相互独立,单位时间的平均到达(上网)率为 λ ;
- (2) n 个通信端口的使用是随机独立的,即任一用户可使用空闲的任一端口,单位时间的平均服务率(上网人数)为 μ ;
- (3) 不限制用户每天的上网次数,即顾客接受一次服务后仍回到顾客总体;
- (4) 学校对用户一般要收取一定数量的线路基本费,在模型中不考虑此费用;
- (5) 学校的收费不是以营利为目的,完全是为了调节线路,控制上网时间,为此,不需要追求经济利益.

14.6.3 模型的建立与求解

由上面的分析,假设用户平均上网的人数(即顾客的平均到达率),服从于参数为 λ 的泊松分布,平均服务(上网)时间服从于参数为 μ 的负指数分布,故问题的排队模型为 $M/M/n/n/\infty$.

问题(1): 已知每个用户平均每天上网1.5 h,每天的总上网时间为 $T = 1.5m$ (h),一天按16 h计算.根据题意要在基本满足需要的条件下,为节省费用,通信端口数应尽可能的少为好,为此,设想让在所有端口满负荷运转的条件下,则每天平均每个通信端口的占用时间应为 $\frac{T}{n} = \frac{1.5m}{n} = 16$ (h),故 $\frac{n}{m} = \frac{1.5}{16} = \frac{1}{10.7}$,即通信端口数 n 与用户数 m 的比例为1:10.7,可近似为1:10.这与实际中通常采用1:10的比例是相符的.

问题(2): 由问题(1)的结果,当 $m = 150$ 时,通信端口数 $n = 16$.由假设1,

用户的平均上网率为 $\lambda = \frac{150}{16} = \frac{75}{8}$ (人/h). 由假设(2), 各端口的平均服务率为 μ , 即每个用户的平均上网时间为 $t = \frac{1}{\mu}$ (h).

假设系统的状态为 k (即由 k 个用户在网上) 的概率为 P_k ($k = 0, 1, 2, \dots, n$), 状态转移率为 $k\mu P_k$, 故得状态的平衡方程为

$$\begin{cases} \mu P_1 = \lambda P_0, \\ \lambda P_{k-1} = k\mu P_k \quad (1 < k < n), \\ \lambda P_{n-1} = n\mu P_n, \end{cases}$$

其中 $\sum_{k=1}^n P_k = 1$. 求解此差分方程, 可得各状态概率为

$$P_k = \frac{1}{k!} \left(\frac{\lambda}{\mu} \right)^k P_0, \quad 0 < k \leq n, \text{ 其中 } P_0 = \left[\sum_{j=0}^n \frac{1}{j!} \left(\frac{\lambda}{\mu} \right)^j \right]^{-1}.$$

而且, 通信端口的平均使用数为

$$L_s = n \left(\frac{\lambda}{\mu} \right) (1 - P_n).$$

系统满员的概率为 $P_n = \frac{1}{n!} \left(\frac{\lambda}{\mu} \right)^n P_0$.

于是, 通信端口有空闲用户能上网的概率为 $\bar{P} = \sum_{k=0}^{n-1} P_k = 1 - P_n$.

当 $\lambda = \frac{75}{8}$ (人/h), $n = 16$, 平均每天单个用户上网 1 h、1.5 h、2 h、3 h、4 h、5 h, 即 $\mu = 1, \frac{2}{3}, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}$ 时, 用户能上网的概率 \bar{P} 、因线路忙用户想上网而上不去产生抱怨的概率 P_{16} 和单位时间端口的平均使用率 L_s 计算结果如表 14-2.

表 14-2 问题(2)的计算结果

| μ | P_{16} | \bar{P} | L_s |
|-------|--------------|-----------|----------|
| 1 | 0.014 666 71 | 0.985 333 | 9.237 5 |
| 2/3 | 0.116 352 | 0.883 648 | 12.426 3 |
| 1/2 | 0.257 403 | 0.742 597 | 13.923 7 |
| 1/3 | 0.467 174 | 0.532 826 | 14.985 7 |
| 1/4 | 0.590 668 | 0.409 332 | 15.35 |
| 1/5 | 0.668 793 | 0.331 207 | 15.525 3 |

问题(3): 根据问题的要求, 采用分段计时收费方案, 可分为“免费→半费→全费→2倍→3倍→4倍……”等时段, 为此, 首先应确定免费时段. 按 $m = 150$

人, $n = 16$ 口, 要保证平均每天每端口为 $150/16 = 9.375$ 人次, 平均上网时间大约为 1.7 h, 即 1 小时 42 分钟. 为了保证一定的可靠性, 同时考虑到问题(1)中端口设计是按 1.5 h 设计的, 于是不妨可确定免费上网时间为 1.5 h, 相应的抱怨概率为 $P_{16}(1.5) = 0.116\ 352$, 全天所有线路饱和的时间为 1.7 h, 且满员的概率为 $P_{16}(1.7) = 0.173\ 387$. 随着 $t = 1/\mu$ 的增加, $P_{16}(t)$ 也增加, 为此, 就按 $P_{16}(t)$ 随时间 t 对 $P_{16}(1.7)$ 增加的倍数来确定对应的上网时间段, 即为分段加倍收费的时间.

事实上, 由 $P_{16}(t) = \frac{P_0}{16!} \left(\frac{\lambda}{\mu} \right)^{16}$ 和 $P_0 = \left[\sum_{j=0}^{16} \frac{1}{j!} \left(\frac{\lambda}{\mu} \right)^j \right]^{-1}$ 可以求出 $t = \frac{1}{\mu}$.

为此根据问题(2)中的计算结果作数据拟合可得到 $P_{16}(t)$ 的近似表达式, 于是有分段计费的时间段为

$$2P_{16}(1.7) = 0.346\ 774, \text{得 } t \approx 2.383\ 7 \approx 2.4;$$

$$3P_{16}(1.7) = 0.520\ 161, \text{得 } t \approx 3.350\ 39 \approx 3.4;$$

$$4P_{16}(1.7) = 0.693\ 548, \text{得 } t \approx 5.237\ 68 \approx 5.2;$$

$$5P_{16}(1.7) = 0.866\ 935, \text{得 } t \approx 6.101\ 64 \approx 6.1.$$

于是, 可以得到分段收费方案:

当 $t < 1.5$ 时, 免费; 当 $1.5 \leq t < 1.7$ 时, 收半费 $\frac{d}{2}$; 当 $1.7 \leq t < 2.4$ 时, 收全费 d ; 当 $2.4 \leq t < 3.4$ 时, 2 倍收费 $2d$; 当 $3.4 \leq t < 5.2$ 时, 3 倍收费 $3d$; 当 $5.2 \leq t < 6.1$ 时, 4 倍收费 $4d$; 当 $t \geq 6.1$ 时, 依次类推, 其中 d 根据学校的实际情况确定.

14.6.4 两点说明

(1) 实际中用户上网数量的多少一定与时间有关系, 早晨、上午、中午、下午、晚上拨号上网的人数肯定是不均衡的. 为此, 应在收费方案中考虑上网的时间因素, 按照“忙时多收, 闲时少收”的原则, 将一天分成早晨、上午、中午、下午、晚上五个时段, 分别赋不同的权值 $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$, 在上面的分段计时收费的同时, 还应考虑上网时段的权值, 综合给出一种“分段计时加权”收费方案, 这种方案可以更好、更有效地起到对通信端口的调节作用;

(2) 该问题是根据本校的信息网建设实际提出来的, 该模型为网络的设计提供了一定的理论依据, 实践证明该模型是符合实际的, 结论是正确的, 并有一定的应用和推广价值.

14.7 参考案例与参考文献

1. 参考案例

(1) 电话总机的设置问题——文献[1]: 163—171

(2) 公交车的调度问题——文献[2]:218—225

(3) 飞机排队问题——文献[3]:112—121

(4) 排队网络问题——文献[4]:194—206

2. 参考文献

[1] 雷功炎. 数学模型讲义. 北京:北京大学出版社,1999

[2] 刘承平. 数学建模方法. 北京:高等教育出版社,2002

[3] 叶其孝. 数学建模教育与国际数学建模竞赛. 合肥:《工科数学》杂志社,1994

[4] 盛友招. 排队论及其在计算机通信中的应用. 北京:北京邮电大学出版社,1998

[5] 陈传贵. 排队论. 北京:北京邮电学院出版社,1994

[6] 朱德通. 最优化模型与实验. 上海:同济大学出版社,2003

[7] 孙荣恒等. 排队论基础. 北京:科学出版社,2002

[8] 刁在筠等. 运筹学. 北京:高等教育出版社,2001

[9] 《运筹学》教材编写组. 运筹学. 修订版. 北京:清华大学出版社,1990

[10] 韩中庚. 校园信息网的设计和调节收费方案. 数学的实践与认识, 2001,31(6):658—659

第十五章 对策论方法

对策论是研究具有斗争或竞争性质现象的一种数学理论和方法,它是运筹学的一个重要分支,它所研究的典型问题是两个或两个以上的参加者在某种对抗性或竞争性的场合下各自做出决策,使自己的一方得到最有利的结果.实际中,在政治、经济、军事活动以及日常生活中都有广泛的应用.

15.1 对策论的基本概念

(1) 对策行为与对策模型

具有竞争或对抗性质的行为称为**对策行为**.具有对策行为的模型称为**对策模型**或**对策**.

实际中的对策模型是多种多样的,根据对策模型的类型可以分为矩阵对策、连续对策、微分对策、阵地对策、随机对策等.

(2) 局中人

一个对策行为(或一局对策)中有权决定自己行动方案的参加者称为**局中人**,用 I 表示局中人的集合.如果有 $n(n > 2)$ 个局中人,则记 $I = \{1, 2, \dots, n\}$.

在对策论中,总是假设每个局中人都是“理智的”,即竞争者是公平理智的竞争.

(3) 策略与策略集

在一局对策中,可供局中人选择的一个实际可行的行动方案称为一个**策略**,每一个局中人都有一组策略,称为**策略集**,记第 i 个局中人的策略集为 $S_i (i \in I)$.

(4) 赢得函数(支付函数)

在一局对策中,每个局中人都选定一个策略形成一个策略组,称为一个**局势**.如果第 i 个局中人的一个策略为 $s_i \in S_i$,则 n 个局中人的策略组为 $s = (s_1, s_2, \dots, s_n)$,是一个局势.全体局势组成的集合 S 表示为每个局中人的策略集的笛卡儿积,即 $S = S_1 \times S_2 \times \dots \times S_n$.

对任意局势 $s \in S$,每个局中人都可以得到一个赢得,如第 i 个局中人的赢得记 $H_i(s)$,则称为第 i 个局中人的**赢得函数**, $H_i(s)$ 可以为正,也可以为负.

由局中人、策略集和赢得函数也就完全确定了对策模型.因此,局中人、策略

集、赢得函数称为对策的三要素.

(5) 对策的分类

对策依据不同的原则可有不同分类:

根据对策的局中人的数目可分为二人对策和多人对策;

多人对策可分为结盟对策和不结盟对策;

根据局中人策略集中的策略数可分为有限对策和无限对策;

根据各局中人的赢得函数的代数和是否为零可分为零和对策和非零和对策;

根据策略与时间的关系可分为静态对策和动态对策;

根据对策的数学模型的类型可分矩阵对策、连续对策、微分对策、阵地对策、随机对策等.

15.2 矩阵对策的概念和理论

15.2.1 矩阵对策的基本概念

如果在一局对策中包含有两个局中人,二局中人都只有有限个策略可供选择,在任一个局势下,两个局中人的赢得之和总是等于0,则称此对策为矩阵对策,又称为二人有限零和对策.

1. 矩阵对策的模型

如果用 I、II 分别表示两局中人,局中人 I 有 m 个纯策略 $\alpha_1, \alpha_2, \dots, \alpha_m$,局中人 II 有 n 个纯策略 $\beta_1, \beta_2, \dots, \beta_n$,则局中人 I、II 的策略集分别为 $S_1 = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$, $S_2 = \{\beta_1, \beta_2, \dots, \beta_n\}$. 当局中人 I、II 分别选择纯策略 α_i, β_j 时,形成了一个纯局势 $(\alpha_i, \beta_j) \in S = S_1 \times S_2$,则对任一个 $(\alpha_i, \beta_j) \in S$,记局中人 I 的赢得值为 a_{ij} ,即局中人 II 赢得值为 $-a_{ij}$ ($i=1, 2, \dots, m; j=1, 2, \dots, n$),故有

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix},$$

称为局中人 I 的赢得矩阵(或为局中人 II 的支付矩阵),即局中人 II 的赢得矩阵为 $-A$.

如果局中人 I、II 的策略集分别为 S_1, S_2 ,局中人 I 的赢得矩阵为 A ,则此矩阵对策就可以确定,其模型记为 $G = \{I, II; S_1, S_2; A\}$,或 $G = \{S_1, S_2; A\}$.

2. 最优策略

定义 15.1 设 $G = \{S_1, S_2; A\}$ 为矩阵对策, $S_1 = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$, $S_2 = \{\beta_1,$

$\beta_2, \dots, \beta_n\}$, $A = (a_{ij})_{m \times n}$, 如果等式

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij} = a_{i^*, j^*}$$

成立, 记 $V_G = a_{i^*, j^*}$, 则称 $V_G = a_{i^*, j^*}$ 为矩阵对策 G 的值, 相应的纯局势 $(\alpha_{i^*}, \beta_{j^*})$ 为对策 G 在纯策略下的解, α_{i^*} 与 β_{j^*} 分别称为局中人 I 与 II 的最优纯策略。

定理 15.1 (G 有解的充要条件) 矩阵对策 $G = \{S_1, S_2; A\}$ 在纯策略意义下有解的充要条件是: 存在纯局势 $(\alpha_{i^*}, \beta_{j^*})$ 使得对一切 $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ 均有

$$a_{ij^*} \leq a_{i^*, j^*} \leq a_{i^*, j}$$

例 15.1 求解矩阵对策

$$G = \{S_1, S_2; A\},$$

其中

$$A = \begin{bmatrix} 2 & 7 & 2 & 1 \\ 2 & 2 & 3 & 4 \\ 3 & 5 & 4 & 4 \\ 2 & 3 & 1 & 6 \end{bmatrix}$$

解 根据矩阵 A 有表 15-1。

表 15-1 局中人的策略及赢得值

| | β_1 | β_2 | β_3 | β_4 | $\min_j a_{ij}$ |
|-----------------|-----------|-----------|-----------|-----------|-----------------|
| α_1 | 2 | 7 | 2 | 1 | 1 |
| α_2 | 2 | 2 | 3 | 4 | 2 |
| α_3 | 3 | 5 | 4 | 4 | 3 |
| α_4 | 2 | 3 | 1 | 6 | 1 |
| $\max_i a_{ij}$ | 3 | 7 | 4 | 6 | |

于是

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij} = a_{31} = 3,$$

则 $V_G = 3$, 对策 G 的解为 α_3, β_1 , 即分别是局中人 I、II 的最优纯策略。

由定理 15.1 和例 15.1 可得到一个事实: 对策 G 的值 $V_G = a_{i^*, j^*}$ 是矩阵 A 中 a_{i^*, j^*} 所在行的最小元素, 又是所在列的最大元素, 即

$$a_{ij^*} \leq a_{i^*, j^*} \leq a_{i^*, j}$$

3. 对策的鞍点

定义 15.2 设对策矩阵为 $A = (a_{ij})_{m \times n}$, 如果存在 (i^*, j^*) 对任意 $i (1 \leq i \leq m)$ 和 $j (1 \leq j \leq n)$ 有 $a_{ij^*} \leq a_{i^*, j^*} \leq a_{i^*, j}$, 则称 (i^*, j^*) 为矩阵 $A = (a_{ij})_{m \times n}$ 的一个鞍

点(或称为对策 G 的鞍点).

由此可将定理 15.1 叙述为:矩阵对策 G 在纯策略意义下有解 $V_G = a_{i^*, j^*}$ 的充要条件是存在 (i^*, j^*) 为 G 的一个鞍点.

对策鞍点的实际意义:对策的值 $V_G = a_{i^*, j^*}$ 是矩阵 A 中第 i^* 行的最小值,第 j^* 列的最大值, (i^*, j^*) 为一个鞍点,相应的 $(\alpha_{i^*}, \beta_{j^*})$ 是对策的一个解,是一个平衡局势,即对策的赢得等于对策的值.也就是说当局中人 I 选择最优策略 α_{i^*} 时,局中人 II 偏离其最优策略 β_{j^*} ,则局中人 I 的赢得会增加,至少不会减少,即局中人 II 的支付会增加,除非他选择 β_{j^*} ,才使支付最少.

同样的,当局中人 II 选择最优策略 β_{j^*} 时,局中人 I 偏离其最优策略 α_{i^*} ,则局中人 I 的赢得会减少,至少不会增加,即局中人 II 的支付会减少,除非局中人 I 选择 α_{i^*} ,才使赢得尽可能的大.

在矩阵对策中,只能在局势 $(\alpha_{i^*}, \beta_{j^*})$ 下,才是双方最理智的选择,使竞争达到一个平衡状态.

4. 矩阵对策解的性质

无差别性: 如果 $(\alpha_{i_1}, \beta_{j_1})$ 和 $(\alpha_{i_2}, \beta_{j_2})$ 都是对策 G 的解,则 $a_{i_1 j_1} = a_{i_2 j_2}$.

可交换性: 如果 $(\alpha_{i_1}, \beta_{j_1})$ 和 $(\alpha_{i_2}, \beta_{j_2})$ 都是对策 G 的解,则 $(\alpha_{i_1}, \beta_{j_2})$ 和 $(\alpha_{i_2}, \beta_{j_1})$ 也是 G 的解.

15.2.2 矩阵对策的混合策略

对于矩阵对策 $G = \{S_1, S_2; A\}$,局中人 I 有把握的赢得至少为 $v_1 = \max_i \min_j a_{ij}$,局中人 II 有把握的支付至多为 $v_2 = \min_j \max_i a_{ij}$. 实际中,一般情况 $v_1 \leq v_2$,特别地,当 $v_1 = v_2$ 时,则称之为对策 G 在纯策略意义下的解,即 $V_G = v_1 = v_2$. 即双方都采取了最理智的策略,为 G 的一个平衡解,局中人 I 的所得正是局中人 II 的所失. 实际中的对策问题未必都是如此,对局中人 I 来讲会选择策略尽力争取得到更多的赢得,而局中人 II 会采取相应的策略尽量保证最小的支付,为了各自的目的,局中人 I 所有可能的策略 $\alpha_1, \alpha_2, \dots, \alpha_m$ 都有可能被采用,而且有一定的随机性,局中人 II 相应可能采用所有可能的策略 $\beta_1, \beta_2, \dots, \beta_n$,与之相对. 局中人 I 和 II 在自己的策略集 S_1 和 S_2 中采取每一策略都有一定的可能性,即按一定的概率来确定相应的各个策略,则构成了混合策略.

1. 混合策略

定义 15.3 设矩阵对策 $G = \{S_1, S_2; A\}$, $S_1 = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$, $S_2 = \{\beta_1, \beta_2, \dots, \beta_n\}$, $A = (a_{ij})_{m \times n}$, 记

$$S_1^* = \left\{ x \in E^m \mid x_i \geq 0, 1 \leq i \leq m, \sum_{i=1}^m x_i = 1 \right\},$$

$$S_2^* = \left\{ y \in E^n \mid y_j \geq 0, 1 \leq j \leq n, \sum_{j=1}^n y_j = 1 \right\},$$

则称 S_1^* 和 S_2^* 分别为局中人 I 和 II 的混合策略集(或策略集). 对每一个 $x \in S_1^*$ 和 $y \in S_2^*$ 分别称为局中人 I 和 II 的混合策略(或策略). (x, y) 称为一个混合局势(或局势). 并称

$$E(x, y) = x^T A y = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_i y_j$$

为局中人的赢得函数. 于是, 构成一个新的对策 $G^* = \{S_1^*, S_2^*; E\}$, 则称 G^* 为 G 的混合扩充.

事实上, 对任意 $x = (x_1, x_2, \dots, x_m)^T \in S_1^*$ 是局中人 I 的一个混合策略, 即意味着: 局中人 I 以概率 x_i 选用策略 α_i , 对任意 $y = (y_1, y_2, \dots, y_n)^T \in S_2^*$ 是局中人 II 的一个混合策略, 即意味着: 局中人 II 以概率 y_j 选用策略 β_j , 并且局中人 I 的赢得为 a_{ij} 的概率为 $x_i y_j$, 则局中人 I 采用混合策略 x 的期望赢得为 $E(x, y) = x^T A y$, 且希望这值越大越好, 至少应是 $\min_{y \in S_2^*} E(x, y)$ (即是局中人 I 最不利, 局

中人 II 最有利的情况). 局中人 I 应选取 $x \in S_1^*$ 使得在最不利中取有利的情形, 即保证自己的赢得不少于 $v_1 = \max_{x \in S_1^*} \min_{y \in S_2^*} E(x, y)$, 对局中人 II 来讲也力争保证自己的所失(支付)至多是 $v_2 = \min_{y \in S_2^*} \max_{x \in S_1^*} E(x, y)$, 显然可有 $v_1 \leq v_2$.

2. 最优策略

定义 15.4 设 $G^* = \{S_1^*, S_2^*; E\}$ 是矩阵对策 $G = \{S_1, S_2; A\}$ 的混合扩充.

如果

$$\max_{x \in S_1^*} \min_{y \in S_2^*} E(x, y) = \min_{y \in S_2^*} \max_{x \in S_1^*} E(x, y),$$

记其值为 V_G , 则称 V_G 为对策 G^* 的值, 相应的混合局势 (x^*, y^*) 为在混合策略意义下的解, x^*, y^* 分别称为局中人 I 和 II 的最优混合策略(最优策略).

注: 当局中人 I 取纯策略 α_k 时, 等价于混合策略

$$x = (x_1, x_2, \dots, x_m)^T \in S_1^*, \quad x_i = \begin{cases} 1, & i = k, \\ 0, & i \neq k, \end{cases}$$

即纯策略是混合策略的特例. 于是, 规定 $G = \{S_1, S_2; A\}$ 与 $G^* = \{S_1^*, S_2^*; E\}$ 不加区别, 用 $G = \{S_1, S_2; A\}$ 表示, 当 G 在纯策略意义下的解不存在时, 自然认为是混合策略意义下的解, 相应的赢得函数为 $E(x, y)$, 因此可得混合策略意义下有解的充要条件.

定理 15.2 对策 $G = \{S_1, S_2; A\}$ 在混合策略意义下有解的充要条件是存在 $x^* \in S_1^*$ 和 $y^* \in S_2^*$ 使 (x^*, y^*) 为 $E(x, y)$ 的一个鞍点. 即对任意 $x \in S_1^*$ 和 $y \in S_2^*$ 有

$$E(x, y^*) \leq E(x^*, y^*) \leq E(x^*, y).$$

15.2.3 矩阵对策的基本定理^[1]

设当局中人 I 取纯策略 α_i 时, 则相应的赢得函数为 $E(i, y)$, 即 $E(i, y) = \sum_{j=1}^n a_{ij}y_j$. 当局中人 II 取纯策略 β_j 时, 则相应的赢得函数为 $E(x, j) = \sum_{i=1}^m a_{ij}x_i$. 于是

$$E(x, y) = \sum_{i=1}^m \sum_{j=1}^n a_{ij}x_iy_j = \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij}y_j \right) x_i = \sum_{i=1}^m E(i, y)x_i.$$

$$\text{同理 } E(x, y) = \sum_{j=1}^n E(x, j)y_j.$$

定理 15.3 设 $x^* \in S_1^*, y^* \in S_2^*$, 则 (x^*, y^*) 是对策 G 的解的充要条件是: 对任意 $i=1, 2, \dots, m$ 和 $j=1, 2, \dots, n$ 有

$$E(i, y^*) \leq E(x^*, y^*) \leq E(x^*, j).$$

定理 15.4 设 $x^* \in S_1^*, y^* \in S_2^*$, 则 (x^*, y^*) 是对策 G 的解的充要条件是: 存在数 v 使得 x^*, y^* 分别是不等式组

$$\begin{cases} \sum_{i=1}^m a_{ij}x_i \geq v \quad (j=1, 2, \dots, n), \\ \sum_{i=1}^m x_i = 1, \\ x_i \geq 0, \quad i=1, 2, \dots, m \end{cases} \quad \text{和} \quad \begin{cases} \sum_{j=1}^n a_{ij}y_j \leq v \quad (i=1, 2, \dots, m), \\ \sum_{j=1}^n y_j = 1, \\ y_j \geq 0, \quad j=1, 2, \dots, n \end{cases}$$

的解, 且 $v = V_G$.

定理 15.5 对任一矩阵对策 $G = \{S_1, S_2; A\}$ 一定存在混合策略意义上的解.

定义 15.5 设 $G = \{S_1, S_2; A\}$, $S_1 = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$, $S_2 = \{\beta_1, \beta_2, \dots, \beta_n\}$, $A = (a_{ij})$, 如果对任意 $j=1, 2, \dots, n$ 有 $a_{i_0j} \geq a_{ij}$, 则称局中人 I 的纯策略 α_{i_0} 优越于纯策略 α_{k_0} .

类似地, 可以定义局中人 II 的纯策略 β_{j_0} 优越于 β_{l_0} .

定理 15.6 设 $G = \{S_1, S_2; A\}$, 如果纯策略 α_1 被其余的纯策略 $\alpha_2, \dots, \alpha_m$ 中之一所优越, 由 G 可得一个新的矩阵对策 $G' = \{S'_1, S_2; A'\}$, 其中 $S'_1 = \{\alpha_2, \dots, \alpha_m\}$, $A' = (a'_{ij})_{(m-1) \times n}$, $a'_{ij} = a_{ij} (i=2, 3, \dots, m; j=1, 2, \dots, n)$, 则有

(1) $V_{G'} = V_G$;

(2) G' 中局中人 II 的最优策略就是在 G 中的最优策略;

(3) 若 $(x_2^*, x_3^*, \dots, x_m^*)^T$ 是 G' 中局中人 I 的最优策略, 则 $x^* = (0, x_2^*, x_3^*, \dots, x_m^*)^T$ 是在 G 中的最优策略.

15.3 矩阵对策的解法

15.3.1 线性方程组的方法

设有矩阵对策 $G = \{S_1, S_2; A\}$, 由定理 15.5 一定存在混合策略意义下的解 (x^*, y^*) , 又由定理 15.4 得 (x^*, y^*) 为 G 的解的充要条件是: 存在 v 使 x^*, y^* 分别为

$$\begin{cases} \sum_{i=1}^m a_{ij}x_i \geq v \quad (j = 1, 2, \dots, n), \\ \sum_{i=1}^m x_i = 1, x_i \geq 0, i = 1, 2, \dots, m \end{cases} \quad \text{和} \quad \begin{cases} \sum_{j=1}^n a_{ij}y_j \leq v \quad (i = 1, 2, \dots, m), \\ \sum_{j=1}^n y_j = 1, y_j \geq 0, j = 1, 2, \dots, n \end{cases}$$

的解, 且 $v = V_G$. 而且可以证明: 当 $x_i^* > 0, y_j^* > 0$ 时, 则 x^*, y^* 分别为

$$\begin{cases} \sum_{i=1}^m a_{ij}x_i = v, j = 1, 2, \dots, n, \\ \sum_{i=1}^m x_i = 1, x_i \geq 0, i = 1, 2, \dots, m \end{cases} \quad \text{和} \quad \begin{cases} \sum_{j=1}^n a_{ij}y_j = v, i = 1, 2, \dots, m, \\ \sum_{j=1}^n y_j = 1, y_j \geq 0, j = 1, 2, \dots, n \end{cases}$$

的解.

15.3.2 线性规划的方法

设有矩阵对策 $G = \{S_1, S_2; A\}$, 由定理 15.2 可以证明, 求 G 在混合策略意义下的解等价于求两个相互对偶的线性规划问题:

$$\begin{aligned} \max w, \\ \text{(P)} \quad \begin{cases} \sum_{i=1}^m a_{ij}x_i \geq w \quad (j = 1, 2, \dots, n), \\ \sum_{i=1}^m x_i = 1, \\ x_i \geq 0, i = 1, 2, \dots, m, \end{cases} \end{aligned} \quad \begin{aligned} \min v, \\ \text{(D)} \quad \begin{cases} \sum_{j=1}^n a_{ij}y_j \leq v \quad (i = 1, 2, \dots, m), \\ \sum_{j=1}^n y_j = 1, \\ y_j \geq 0, j = 1, 2, \dots, n \end{cases} \end{aligned}$$

的解, 不妨设 $v > 0$, 令 $x'_i = \frac{x_i}{v}$ ($i = 1, 2, \dots, m$), $y'_j = \frac{y_j}{v}$ ($j = 1, 2, \dots, n$) 分别代入问题 (P) 和 (D) 中, 线性规划问题 (P) 和 (D) 变为

$$\min z = \sum_{i=1}^m x'_i, \quad \max w = \sum_{j=1}^n y'_j,$$

$$(P') \begin{cases} \sum_{i=1}^m a_{ij} x'_i \geq 1, j = 1, 2, \dots, n, \\ x'_i \geq 0, i = 1, 2, \dots, m, \end{cases} \quad (D') \begin{cases} \sum_{j=1}^n a_{ij} y'_j \leq 1, i = 1, 2, \dots, m, \\ y'_j \geq 0, j = 1, 2, \dots, n, \end{cases}$$

则问题(P')和(D')也是相互对偶的线性规划问题,可以利用单纯形法求解,可以得到 \mathbf{x}'^* 和 \mathbf{y}'^* ,通过反变换 $x_i = x'_i v, y_j = y'_j v$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$),即可得到 \mathbf{x}^* 和 \mathbf{y}^* 为矩阵对策的解和对策的值 V_G .

15.4 双矩阵对策

在矩阵对策中,局中人I的所得就是局中人II的所失,对策结果可用一个矩阵表示.而在非零和的对策中就不同了,若局中人I选择策略 $\alpha_i \in S_1$,而局中人II选择策略 $\beta_j \in S_2$,则对策局势为 $(\alpha_i, \beta_j) \in S$,相应的局中人I的赢得为 a_{ij} ,局中人II的赢得不再是 $-a_{ij}$,而是 b_{ij} ,即对策结果为 (a_{ij}, b_{ij}) .这种对策通常记为 $G = \{S_1, S_2; \mathbf{A}, \mathbf{B}\}$,其中 $\mathbf{A} = (a_{ij}), \mathbf{B} = (b_{ij})$,分别是局中人I和II的赢得矩阵,故称为二人有限非零和对策,或双矩阵对策.

在非零和矩阵对策中,二局中人并不是完全对立的,即局中人I的所得不再是局中人II的所失,因此二局中人既可以合作,也可以不合作.在不合作时,假设二局中人之间不能互通信息,也没有任何形式的联合或协商,即双方是直接对抗的.在合作的时候,对策双方则可能有共同的认识,譬如,双方都认为某种结果比其他的结果对自己有利.下面分两种情况讨论.

15.4.1 非合作的双矩阵对策

1. 引例

囚犯难题(20世纪50年代,德雷歇(Dresher)和弗拉德(Flood)):

设有两个人因藏被盗物品而被捕,现分别关押受审.二人都明白,如果都拒不承认,现有的证据不足以证明他们偷盗,而只能以窝赃罪判处一年监禁;两人要是都承认了将各判9年;但如果一人招认而另一人拒不承认,那么坦白者将会从宽处理获得释放,而抗拒者从严被判10年.这两个囚犯该选择什么策略?是坦白交待,还是拒不承认呢?

假设囚犯I与II的第一个策略都是坦白认罪,第二个策略则是拒不交待,对他们判处监禁的年数表示他们的赢得,则他们的赢得矩阵为

II :

$$\begin{array}{cc} & \beta_1 & \beta_2 \\ \text{I : } & \alpha_1 \begin{bmatrix} (-9, -9) & (0, -10) \\ (-10, 0) & (-1, -1) \end{bmatrix} & \alpha_2 \end{array}$$

在此对策中,二囚犯是隔离受审,因此,他们不能合作,只有各自为自己的前途考虑,总是被监禁的年数越少越好,故他们的最优策略均为坦白交待,且对策值对各自来讲为 $v = -9$,但实际上 $(-9, -9)$ 对二人来说都不是最好的,相比之下结果 $(-1, -1)$ 更好. 这个问题之所以称为难题,主要体现在两方面:

难题(一): 二局中人应该选什么作为目标? 他们作为独立的个体,同时又是集体中的一员,应该怎样做最好? 即在个体的合理性和集体的合理性之间有冲突.

难题(二): 把这个问题看成一次性对策还是可以重复进行下去的多次对策? 即是一次审讯还是多次审讯? 若是一次审讯,当然是坦白好,因为没有理由相信另一个囚犯会为你着想. 但是,若可重复审讯下去,结果就会不同了.

这个问题也用于模拟各类带有冲突性的问题,例如裁军、谈判、价格大战等等.

注: 如果把双矩阵对策分解为两个矩阵对策,其中一个只考虑局中人 I 的赢得,另一个只考虑局中人 II 的赢得,则这两个对策总是有解的. 各自按照最大最小原则都可以得到最优策略,其对策值分别为 v_1 和 v_2 ; 但当局中人 I 与 II 各以理性地来参加对策时,对策的结果 (v_1, v_2) 不一定是最佳的结果. 上面的例子也说明了这一点,为此,对于非零和对策问题,不能用零和对策的方法来求解. 下面引入新的平衡局势和解的概念.

2. 纳什(Nash)平衡点

在非零和对策中,对任意的 $x = (x_1, x_2, \dots, x_m)^T \in S_1, y = (y_1, y_2, \dots, y_n)^T \in S_2$, 定义

$$\begin{aligned} E_1(x, y) &= x^T A y = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_i y_j, \\ E_2(x, y) &= x^T B y = \sum_{i=1}^m \sum_{j=1}^n b_{ij} x_i y_j \end{aligned}$$

分别表示局中人 I 和 II 的赢得函数.

定义 15.6 如果存在一对策略 $x^* \in S_1, y^* \in S_2$ 使得对任意 $x \in S_1, y \in S_2$ 都有

$$\begin{aligned} E_1(x, y^*) &\leq E_1(x^*, y^*), \\ E_2(x^*, y) &\leq E_2(x^*, y^*) \end{aligned}$$

成立,则称策略对 (x^*, y^*) 为对策的一个平衡点,或称纳什(Nash)平衡点.

纳什 1951 年证明了平衡点的存在性定理.

定理 15.7 (Nash) 任何具有有限个纯策略的二人对策 (包括零和对策与非零和对策) 至少存在一个平衡点.

纳什定理仅说明了平衡点的存在性, 并没有说明平衡点的求解方法, 而且, 平衡点不是惟一的. 一般说来求解平衡点是困难的, 已有的方法计算量很大, 这一问题仍是需要进一步研究的问题之一.

3. 非零和对策解的概念

(1) 纳什意义下的解

定义 15.7 如果对策的每一对平衡点都是可交换的, 即若 (x_1, y_1) 与 (x_2, y_2) 是平衡点, 则 (x_1, y_2) 与 (x_2, y_1) 也是平衡点, 此时称对策存在纳什意义下的解. 其解就是所有可交换的平衡点的集合.

思考题: 零和对策的解是否是纳什意义下的解; 囚犯问题是否存在纳什意义下的解?

(2) 严格意义下的解

首先给出 (x_2, y_2) 优超 (x_1, y_1) 的概念: 如果

$$E_1(x_2, y_2) \geq E_1(x_1, y_1), E_2(x_2, y_2) \geq E_2(x_1, y_1),$$

并且其中至少有一个不等式是严格成立的, 则称 (x_2, y_2) 优超 (x_1, y_1) .

如果 (x, y) 优超所有的策略对 (即没有优超于 (x, y) 的策略对), 则称 (x, y) 是帕雷托 (Pareto) 最优策略对.

定义 15.8 如果对策满足下列条件:

- ① 在帕雷托最优策略对中有平衡点;
- ② 所有帕雷托最优平衡点都是可交换的, 并且有相同的赢得.

则称帕雷托最优平衡点的集合为对策在严格意义下的解.

对于囚犯问题: 因为策略对 $((0, 1), (0, 1))$ 的赢得为 $(-1, -1)$, 它优超于该对策中惟一的平衡点 $((1, 0), (1, 0))$, 但由于 $((0, 1), (0, 1))$ 不是平衡点, 所以囚犯问题没有严格意义下的解.

(3) 完全弱意义下的解

基本思想: 利用优超的概念来减弱上述严格意义下的解的定义. 首先根据局中人 I 的赢得, 从 S_1 中删去那些针对 S_2 却又被其他策略优超的策略, 并把剩下的策略记为 S_1^1 ; 然后根据局中人 II 的赢得从 S_2 中删除那些针对 S_1^1 却又被优超的策略, 把剩下的记为 S_2^1 . 如此继续下去, 可以得到 $S_1^2, S_2^2; S_1^3, S_2^3; \dots$, 直到没有可删除的策略为止, 记最后得到的策略集为 S_1^T, S_2^T .

定义 15.9 若以 S_1^T, S_2^T 为策略集的简化对策在严格意义下是可解的, 则称以 S_1, S_2 为策略集的对策在完全弱意义下是可解的. 而且简化对策在严格意义下的解的集合就是原来对策的解的集合.

对于囚犯问题: $S_1 = \{\alpha_1, \alpha_2\}$, $S_2 = \{\beta_1, \beta_2\}$, 按优超原则可简化为 $S_1^T = \{\alpha_1\}$, $S_2^T = \{\beta_1\}$, 则简化对策为 $G = \{S_1^T, S_2^T; A, B\}$. 二人的赢得为 $(-9, -9)$, 所以平衡点为 (α_1, β_1) , 故简化对策在严格意义下有解, 即囚犯问题在完全弱意义下有解, 其解为 (α_1, β_1) , 对策的值为 $(-9, -9)$.

综上给出了三种解的定义, 共同的不足之处, 是不能使所有的对策问题在同一个解的意义下都有解. 实际中还有一些其他形式的解的定义, 虽然能使所有对策问题可解, 但平衡点的数目却不能显著减少, 因而使计算难度不减, 这些仍是需要研究的课题.

15.4.2 合作的双矩阵对策

这里所说的合作是指在对策之前, 二人就对策问题的协商, 互通信息, 确定双方的联合策略, 即一致同意局中人 I 采用策略 $x \in S_1$, 局中人 II 采用策略 $y \in S_2$. 然而, 这种协商 (即合作) 应该对双方有一定的约束作用. 但事实上, 实际中可能有某个人出现违约的情况. 这里要研究的问题是: 当二局中人能够合作时, 他们可能采取什么策略? 这只要确定赢得区域.

在双矩阵对策中, 因为双方的赢得函数为

$$E_1(x, y) = \sum_{i=1}^m \sum_{j=1}^n a_{ij} x_i y_j \text{ 和 } E_2(x, y) = \sum_{i=1}^m \sum_{j=1}^n b_{ij} x_i y_j,$$

所谓的赢得区域就是双方对任意的 $x \in S_1, y \in S_2$ 所有可能的赢得值, 即 E_1, E_2 的变化区域.

注意这样一个事实: 合作对策的赢得区域一定比非合作的赢得区域大, 这是因为在合作对策中, 可以采用联合的随机策略, 所有可能的策略要比非合作的策略多得多. 例如: 如果非合作对策的赢得为 (u_1, v_1) 与 (u_2, v_2) , 那么在合作的情况下, 以概率 λ 采用赢得为 (u_1, v_1) 的策略, 以概率 $1 - \lambda$ 采用赢得为 (u_2, v_2) 的策略, 于是双方的期望赢得为

$$\lambda(u_1, v_1) + (1 - \lambda)(u_2, v_2).$$

由此可知, 合作对策的赢得区域 R 是非合作对策赢得区域的凸闭包, 即把端点在非合作对策赢得区域内的所有线段都包含在内的最小闭区域 R . 其顶点是双方都采用纯策略时的赢得值, 但并非是所有纯策略对的赢得都必定在 R 的顶点上, 即也可能在 R 内部. 据此, 可以很容易确定出赢得区域 R .

更一般的问题需要研究纳什谈判集和纳什谈判定理, 从而可求出纳什谈判解^[3,4,5].

15.5 玫瑰有约问题

15.5.1 问题的提出

目前,许多城市大龄青年的婚姻问题已引起了妇联、工会和社会团体组织的关注.某单位现有 20 对大龄青年男女,每个人的基本条件都不相同,如外貌、性格、气质、事业、财富等.每项条件通常可以分为五个等级 A、B、C、D、E,如外貌、性格、气质、事业可分为很好、好、较好、一般、差;财富可分为很多、多、较多、一般、少.每个人的择偶条件也不尽相同,即对每项基本条件的要求是不同的.该单位拟根据他(她)们的年龄、基本条件和要求条件进行牵线搭桥.下面给出 20 对大龄青年男女的年龄、五项基本条件和要求条件(如表 15-2 和表 15-3).一般认为,男青年至多比女青年大 5 岁,或女青年至多比男青年大 2 岁,并且要至少满足个人要求 5 项条件中的 2 项,才有可能配对成功.该单位希望根据每个人的情况和要求,建立数学模型解决下列问题:

(1) 在尽量满足个人要求的条件下,给出一种最佳的配对方案,并使得配对成功率尽可能的高;

(2) 给出一种 20 对男女青年可同时配对的最佳方案,使得全部配对成功的可能性最大;

(3) 如果 20 对男女青年双方都相互了解了对方的条件和要求,让每个人做出一次选择,只有当男女双方相互选中对方时才认为能够配对成功,每人只有一次选择机会.请问 20 对男女青年应该如何选择,使得自己配对成功的可能性最大?按你的选择方案最多能配对成功多少对?

15.5.2 问题的分析

该问题是现实生活中的实际问题,主要就是确定合理配对方案,使得在尽量满足个人要求的条件下,使配对成功率尽可能高.由于每个人的基本条件和要求都是给定的,双方彼此是知道的,而且相互之间有很大的差异,如果完全按照要求条件来组合配对,则其成功的可能性是很小的.一对青年男女能否配对成功,主要是取决于相互之间的好感的程度——“满意度”,单方面的高满意度也不一定能配对成功.任意一对男女的配对可以看成是一个随机事件,按某一概率可能配对成功,或不成功.在这里双方的满意度主要反映出了一个人对另一个人的客观和主观的看法,因此,满意度的定义成为解决问题的一个关键.

所谓的“成功率”,就是男女双方最终配对成功的概率.实际上,可以用他们相互之间的满意度来间接刻画.相互的满意度越高,双方配对成功的概率就越大.

表 15-2 男青年的基本条件和要求条件

| 男青年 | 基 本 条 件 | | | | | | 要 求 条 件 | | | | |
|----------|---------|----|----|----|----|----|---------|----|----|----|----|
| | 外貌 | 性格 | 气质 | 事业 | 财富 | 年龄 | 外貌 | 性格 | 气质 | 事业 | 财富 |
| B_1 | A | C | B | C | A | 29 | A | A | C | B | D |
| B_2 | C | A | B | A | D | 29 | B | A | B | B | C |
| B_3 | B | B | A | B | B | 28 | B | A | A | B | C |
| B_4 | C | A | B | B | D | 28 | C | A | B | C | D |
| B_5 | D | B | C | A | A | 30 | C | B | B | B | E |
| B_6 | C | B | C | B | B | 28 | B | B | C | D | C |
| B_7 | A | B | B | D | C | 30 | C | B | B | D | C |
| B_8 | B | A | B | C | D | 30 | A | B | C | C | D |
| B_9 | A | D | C | E | B | 28 | A | A | A | C | C |
| B_{10} | D | B | A | A | A | 28 | A | B | A | D | E |
| B_{11} | B | A | C | D | A | 32 | A | B | C | D | B |
| B_{12} | A | B | C | A | B | 29 | B | A | B | B | C |
| B_{13} | B | A | D | E | C | 28 | A | C | B | B | C |
| B_{14} | A | A | B | B | D | 30 | A | C | C | D | C |
| B_{15} | A | B | B | C | C | 28 | A | A | B | C | D |
| B_{16} | D | E | B | A | A | 30 | A | A | A | E | E |
| B_{17} | B | A | B | A | D | 28 | B | A | B | B | C |
| B_{18} | A | B | A | C | B | 31 | B | B | A | C | C |
| B_{19} | C | D | A | A | A | 29 | A | B | A | E | D |
| B_{20} | A | B | C | D | E | 27 | B | C | B | D | B |

表 15-3 女青年的基本条件和要求条件

| 女青年 | 基 本 条 件 | | | | | | 要 求 条 件 | | | | |
|----------|---------|----|----|----|----|----|---------|----|----|----|----|
| | 外貌 | 性格 | 气质 | 事业 | 财富 | 年龄 | 外貌 | 性格 | 气质 | 事业 | 财富 |
| G_1 | A | C | C | D | A | 28 | B | A | B | A | D |
| G_2 | B | A | B | A | D | 25 | C | B | B | A | B |
| G_3 | C | B | A | E | A | 26 | B | A | C | B | C |
| G_4 | A | B | B | C | D | 27 | A | A | B | B | A |
| G_5 | B | D | C | E | C | 25 | A | B | C | B | B |
| G_6 | A | C | B | C | A | 26 | B | A | B | B | C |
| G_7 | D | C | B | A | B | 30 | C | B | A | A | C |
| G_8 | A | B | A | E | C | 31 | B | A | B | A | B |
| G_9 | A | A | A | C | E | 26 | C | B | B | B | A |
| G_{10} | B | C | D | B | B | 27 | B | B | A | A | C |
| G_{11} | A | B | B | C | B | 28 | C | B | A | B | C |
| G_{12} | B | E | C | E | A | 26 | A | A | B | B | E |
| G_{13} | E | A | C | B | B | 26 | C | A | B | C | C |
| G_{14} | B | B | C | A | A | 25 | B | A | A | B | D |
| G_{15} | C | B | A | A | C | 29 | B | A | B | B | B |
| G_{16} | B | A | C | D | C | 28 | B | A | B | B | A |
| G_{17} | A | E | E | D | A | 25 | A | A | D | A | C |
| G_{18} | A | A | B | B | C | 28 | C | A | B | A | C |
| G_{19} | B | A | C | C | E | 25 | B | B | B | A | A |
| G_{20} | D | B | A | C | D | 29 | B | B | A | B | B |

注：表中的要求条件一般是指不低于所给的条件

对于问题(1),要使配对成功率尽可能的高,也就是给出一种方案,使得20对男女的配对后的满意度之和最高.

对于问题(2),要使20对男女青年同时配对,使得全部同时配对成功的可能性(概率)最大.

对于问题(3),因为每个人只能选择一次,能否配对成功主要取决于双方是否选中对方,即要看双方彼此的满意度如何.实际中,假如一个男青年 $B_i (1 \leq i \leq 20)$ 对一个女青年 $G_j (1 \leq j \leq 20)$ 的满意度最高,但 G_j 对 B_i 的满意度不一定最高,即若 B_i 选择 G_j ,但 G_j 不一定选择 B_i . 因此, B_i 与 G_j 不一定能配成对,反之亦然. 现在的问题是谁选谁,使配对成功的可能性最大呢? 这个问题实际上是男女双方在彼此基本了解的情况下,在保证自己一定满意度的条件下做出自己的选择,也需要猜测对方会做出什么的选择. 因此,这个问题可以转化为男女双方的对策问题,即转化为求男女双方的非零和对策的纳什平衡点的问题.

15.5.3 模型的假设与符号说明

1. 模型的假设

- (1) 题目所给出的男女青年的评价是客观真实的;
- (2) 每个人在选择对方时都是理智的;
- (3) 五项条件在选择对方时所起的作用是均等的.

2. 符号的约定

B_i 与 G_j 分别表示男青年与女青年, x_{ij} 表示第 i 个男青年与第 j 个女青年配对时取值1,否则取值0; $a_{ik}^{(0)}$ 与 $a_{ik}^{(1)}$ 分别表示第 i 个女青年与男青年的第 k 项基本条件的量化指标; $b_{ik}^{(0)}$ 与 $b_{ik}^{(1)}$ 分别表示第 i 个女青年与男青年的第 k 项要求条件的量化指标; $S_{ij}^{(0)}(k)$ 表示第 i 个女青年对第 j 个男青年的第 k 项条件的满意度; $S_{ij}^{(1)}(k)$ 表示第 i 个男青年对第 j 个女青年的第 k 项条件的满意度; $S_{ij}^{(0)}$ 表示第 i 个女青年对第 j 个男青年的综合满意度; $S_{ij}^{(1)}$ 表示第 i 个男青年对第 j 个女青年的综合满意度; SS_{ij} 表示第 i 个男青年与第 j 个女青年的综合相互满意度; P_{ij} 表示第 i 个男青年与第 j 个女青年配对成功的概率; P_r 表示所给方案的使配对成功的总概率; $A = (SS_{ij}^{(1)})_{20 \times 20}$, $B = (SS_{ij}^{(0)})_{20 \times 20}$ 表示所有男青年,女青年对异性青年的满意度矩阵; d 表示男(女)青年的要求条件等级距离最高等级 A 的档数; h 表示男(女)青年的基本条件等级高出对方所要求条件等级的档数. 其中 $i = 1, 2, \dots, 20; j = 1, 2, \dots, 20; k = 1, 2, \dots, 5$.

15.5.4 模型的准备

1. 条件的量化处理

对于每个人的外貌、性格、气质、事业和财富五项条件的5个等级A,B,C,D

和 E 分别作量化处理,根据层次分析中关于条件等级差的度量标准,对 A,B,C,D,E 分别赋权值为 9,7,5,3,1. 于是根据表 15-2 和表 15-3 可以得到男女青年的基本条件量化矩阵和要求条件量化矩阵(或称权值矩阵)分别记为

$$\begin{cases} \mathbf{A}^{(1)} = (a_{ik}^{(1)})_{20 \times 5}, \mathbf{A}^{(0)} = (a_{ik}^{(0)})_{20 \times 5}, \\ \mathbf{B}^{(1)} = (B_{ik}^{(1)})_{20 \times 5}, \mathbf{B}^{(0)} = (b_{ik}^{(0)})_{20 \times 5}. \end{cases} \quad (15.1)$$

2. 条件过滤

由于问题是明确要求“男青年的年龄至多比女青年大 5 岁,而女青年的年龄至多比男青年大 2 岁”,以及“至少满足个人要求 5 项条件中的 2 项”. 在 20 对男女青年中所有可能的配对中,首先应将不满足这些基本条件的情况过滤掉. 由(15.1)式用 Matlab 编程进行过滤可以过滤掉 58 对组合.

3. 满意度的确定

(1) 对单项条件的满意度

要确定 G_j 对 B_i 的第 k ($1 \leq i, j \leq 20, 1 \leq k \leq 5$) 项条件的满意度. 首先要注意到两个事实:其一,如果 B_i 的基本条件 $a_{ik}^{(1)}$ 比 G_j 的要求条件 $b_{jk}^{(0)}$ 差的越多,则 G_j 对 B_i 的第 k 项条件的满意度 $S_{ji}^{(0)}(k)$ 就越小,反之亦然. 也就是说,如果一方的实际条件比对方期望(要求)的条件差距越大,则对方对另一方失望就越大,即满意度就越小. 其二,如果 B_i 的基本条件 $a_{ik}^{(1)}$ 比 G_j 的要求条件 $b_{jk}^{(0)}$ 高,则 G_j 对 B_i 的第 k 项条件的满意度 $S_{ji}^{(0)}(k)$ 就会增加,但增加不会太多. 即当一方的实际条件高于对方期望(要求)的条件时,则对方对另一方的好感(相对要求条件)增加不会太大.

根据上面的两个事实,先以女青年 G_j 对男青年 B_i 的第 k 项条件的满意度 $S_{ji}^{(0)}(k)$ 为例. 若 B_i 的实际条件比 G_j 的要求差,那么 G_j 对 B_i 该项指标的满意度将迅速减小,减小的速度一般会与 G_j 的要求条件相差的档数有关,二者会成一定的比例关系. 当 B_i 的实际条件比 G_j 的要求条件差得很大(例如,三个档次以上)时,则认为 G_j 对 B_i “失去信心”,即满意度为 0. 若 B_i 的实际条件比 G_j 的要求条件还要好,那么 G_j 对 B_i 的这项条件的满意度会略有增加,但增加的不会太大.

根据实际情况,一般认为如果男青年 B_i 的条件达到女青年 G_j 的要求条件时,那么, G_j 会以 90% 的可能接受 B_i , 此时即可以认为 G_j 对 B_i 的第 k 项条件的满意度 $S_{ji}^{(0)}(k) = 0.9$.

下面首先给出 G_j 对 B_i 的第 k 项条件的满意度 $S_{ji}^{(0)}(k)$ 的定义:

$$S_{ji}^{(0)}(k) = \begin{cases} 0.9 + \frac{0.1}{d}h, a_{ik}^{(1)} \geq b_{jk}^{(0)}, \\ 0.9 \cdot \frac{a_{ik}^{(1)}}{b_{jk}^{(0)}}, a_{ik}^{(1)} < b_{jk}^{(0)} \text{ 且 } b_{jk}^{(0)} - a_{ik}^{(1)} < 3, \left(\begin{matrix} i, j = 1, 2, \dots, 20; \\ k = 1, 2, \dots, 5 \end{matrix} \right); \\ 0, a_{ik}^{(1)} < b_{jk}^{(0)} \text{ 且 } b_{jk}^{(0)} - a_{ik}^{(1)} \geq 3 \end{cases} \quad (15.2)$$

如果 G_j 的第 k 项条件的要求为 C , 则根据 B_i 的第 k 项条件的不同情况, G_j 对 B_i 的第 k 项条件的满意度 $S_{ji}^{(0)}(k)$ 的取值如图 15-1.

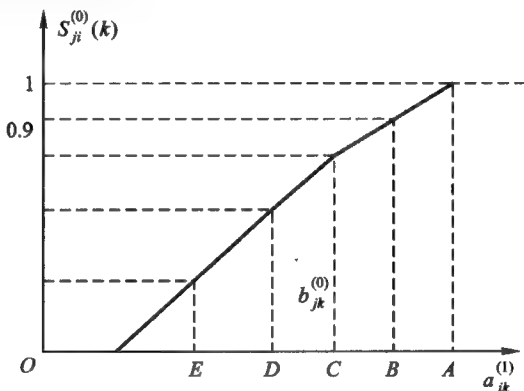


图 15-1 G_j 的要求为 C 时, G_j 对 B_i 的满意度

同理定义 B_i 对 G_j 的第 k 项条件的满意度 $S_{ij}^{(1)}(k)$ 如下:

$$S_{ij}^{(1)}(k) = \begin{cases} 0.9 + \frac{0.1}{d}h, a_{jk}^{(0)} \geq b_{ik}^{(1)}, \\ 0.9 \cdot \frac{a_{jk}^{(0)}}{b_{ik}^{(1)}}, a_{jk}^{(0)} < b_{ik}^{(1)} \text{ 且 } b_{ik}^{(1)} - a_{jk}^{(0)} < 3, \left(\begin{matrix} i, j = 1, 2, \dots, 20; \\ k = 1, 2, \dots, 5 \end{matrix} \right); \\ 0, a_{jk}^{(0)} < b_{ik}^{(1)} \text{ 且 } b_{ik}^{(1)} - a_{jk}^{(0)} \geq 3 \end{cases} \quad (15.3)$$

(2) 综合满意度

一对男女青年相互之间的综合满意度主要取决于对对方的各项条件的满意度 $S_{ij}^{(1)}(k)$ 和 $S_{ji}^{(0)}(k)$ ($1 \leq i, j \leq 20; k = 1, 2, \dots, 5$).

由假设, 可以定义男青年 B_i 对女青年 G_j 的综合满意度为

$$S_{ij}^{(1)} = \frac{1}{5} \sum_{k=1}^5 S_{ij}^{(1)}(k) \quad \left(\begin{matrix} i, j = 1, 2, \dots, 20; \\ k = 1, 2, \dots, 5 \end{matrix} \right). \quad (15.4)$$

由 (15.1), (15.3), (15.4) 式计算可得到结果.

同理, 定义女青年 G_j 对男青年 B_i 的综合满意度为

$$S_{ji}^{(0)} = \frac{1}{5} \sum_{k=1}^5 S_{ji}^{(0)}(k) \quad \left(\begin{array}{l} i, j = 1, 2, \dots, 20; \\ k = 1, 2, \dots, 5 \end{array} \right). \quad (15.5)$$

由(15.1), (15.4), (15.6)式可以计算出结果.

(3) 相互满意度

男青年 B_i 与女青年 G_j 的相互满意度定义为

$$SS_{ij} = \begin{cases} \sqrt{S_{ij}^{(1)} S_{ji}^{(0)}}, & \text{当 } B_i \text{ 与 } G_j \text{ 满足可能配对的条件时,} \\ 0, & \text{当 } B_i \text{ 与 } G_j \text{ 不满足可能配对的条件时.} \end{cases}$$

经过计算可得到男女青年相互之间的满意度.

15.5.5 模型的建立与求解

问题(1): 根据问题的要求, 欲使得在尽量满足个人要求的条件下, 使配对成功率尽可能的高. 事实上, 我们可以用 20 对男女青年的满意度指标 $SS_{ij} (i, j = 1, 2, \dots, 20)$ 之和来刻画总的配对成功的成功率 P_T , 于是我们将问题归结为所有 20 对男女青年如何配对能使得 $\sum_{i=1}^{20} \sum_{j=1}^{20} SS_{ij} x_{ij}$ 有最大值, 即问题的模型为

$$\begin{aligned} \max z = & \sum_{i=1}^{20} \sum_{j=1}^{20} SS_{ij} x_{ij}, \\ \text{s. t. } & \begin{cases} \sum_{i=1}^{20} x_{ij} = 1, j = 1, 2, \dots, 20, \\ \sum_{j=1}^{20} x_{ij} = 1, i = 1, 2, \dots, 20, \\ x_{ij} = 0, \text{ 或 } 1 (i, j = 1, 2, \dots, 20). \end{cases} \end{aligned} \quad (15.6)$$

这是一个 0-1 规划问题, 用匈牙利方法(或 LINGO 软件)求解可得最优配对方案如表 15-4. 最优值(总满意度)为 $z = 16.75132$.

表 15-4 问题(1)的最优配对方案

| | | | | | | | | | | |
|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 男 | B_1 | B_2 | B_3 | B_4 | B_5 | B_6 | B_7 | B_8 | B_9 | B_{10} |
| 女 | G_4 | G_{18} | G_{15} | G_{20} | G_2 | G_5 | G_3 | G_{13} | G_{17} | G_9 |
| 满意度 | 0.849 9 | 0.909 5 | 0.887 4 | 0.801 6 | 0.893 0 | 0.777 7 | 0.834 9 | 0.836 0 | 0.646 5 | 0.932 7 |
| 男 | B_{11} | B_{12} | B_{13} | B_{14} | B_{15} | B_{16} | B_{17} | B_{18} | B_{19} | B_{20} |
| 女 | G_{16} | G_{10} | G_7 | G_1 | G_6 | G_{19} | G_{14} | G_{11} | G_8 | G_{12} |
| 满意度 | 0.829 6 | 0.822 4 | 0.756 8 | 0.929 9 | 0.867 4 | 0.779 6 | 0.897 2 | 0.907 5 | 0.878 4 | 0.713 3 |

问题(2): 要使 20 对男女青年同时配对, 使得全部配对成功的可能性最大.

记 20 对男女青年可同时配对成功的概率为 $P = \prod_{i,j} P_{ij}$, 即问题为求

$$\max P = \prod_{i,j} P_{ij}.$$

由于 $P_{ij} \propto SS_{ij}$, 则问题等价于求一个 20 对男女青年同时配对的方案使目标函数

$$\max z = \prod_{i=1}^{20} \prod_{j=1}^{20} SS_{ij} x_{ij}$$

有最大值, 其约束条件与 (15.6) 式相同.

令 $C_{ij} = \ln(SS_{ij})$, 考虑相应的目标函数

$$\max Z = \sum_{i=1}^{20} \sum_{j=1}^{20} C_{ij} x_{ij},$$

约束条件不变, 同理求解可得最优的配对方案如表 15-5, 其最优值为 $z = 0.0197$.

表 15-5 问题(2)的最优配对方案

| | | | | | | | | | | |
|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 男 | B_1 | B_2 | B_3 | B_4 | B_5 | B_6 | B_7 | B_8 | B_9 | B_{10} |
| 女 | G_2 | G_{18} | G_6 | G_{14} | G_{19} | G_5 | G_3 | G_{13} | G_{11} | G_{20} |
| 满意度 | 0.881 2 | 0.909 5 | 0.861 2 | 0.869 9 | 0.822 3 | 0.777 7 | 0.834 9 | 0.836 0 | 0.757 7 | 0.850 6 |
| 男 | B_{11} | B_{12} | B_{13} | B_{14} | B_{15} | B_{16} | B_{17} | B_{18} | B_{19} | B_{20} |
| 女 | G_1 | G_{17} | G_7 | G_{12} | G_{15} | G_9 | G_{16} | G_8 | G_4 | G_{10} |
| 满意度 | 0.843 5 | 0.733 9 | 0.756 8 | 0.807 5 | 0.837 9 | 0.850 1 | 0.812 0 | 0.839 8 | 0.853 2 | 0.725 9 |

问题(3): 根据问题的分析可以建立该问题的对策模型.

假设有两个虚拟的局中人, 男青年 I 和女青年 II, 其策略集分别为

$$S_1 = \{B_i \mid i = 1, 2, \dots, 20\}; \quad S_2 = \{G_j \mid j = 1, 2, \dots, 20\};$$

指标集为

$$I = \{i \mid i = 1, 2, \dots, 20\}; \quad J = \{j \mid j = 1, 2, \dots, 20\}.$$

局中人 I 和局中人 II 采用纯策略 B_i 和 G_j 的赢得分别为 B_i 对 G_j 的满意度 $SS_{ij}^{(1)}$ 和 G_j 对 B_i 的满意度 $SS_{ji}^{(0)}$ ($i, j = 1, 2, \dots, 20$), 于是可以得到二局中人的赢得矩阵分别为 $A = (SS_{ij}^{(1)})_{20 \times 20}$ 和 $B = (SS_{ji}^{(0)})_{20 \times 20}$. 因此, 我们可以构造出了一个二人非零和双矩阵对策模型

$$\Gamma = (S_1, S_2; A, B).$$

按照每个人的条件量化的原则, 对于每个人来说, 自己在每个异性心目中的地位——满意度的大小都是可以估算出来的, 即对策双方的赢得矩阵是已知的. 对策双方不应该追求自己的最高赢得, 要保证配对成功率, 只有选择对策双方都能接受的策略, 即确保男女双方相互选中对方, 才是最理智的行为. 于是, 根据假设(2), 问题就是求对策 $\Gamma = (S_1, S_2; A, B)$ 的纳什平衡点的问题了.

所谓的纳什平衡点:在对策 $\Gamma = (S_1, S_2; \mathbf{A}, \mathbf{B})$, $\mathbf{A} = (a_{ij})$, $\mathbf{B} = (b_{ij})$ 中,若有策略对 $(B_{i^*}, G_{j^*}) \in S_1 \times S_2$, 即 $(i^*, j^*) \in I \times J$, 使得

$$\begin{cases} a_{ij^*} \leq a_{i^*j^*}, i \in I, \\ b_{i^*j} \leq b_{i^*j^*}, j \in J, \end{cases} \quad (15.7)$$

则称 (i^*, j^*) 为对策 $\Gamma = (S_1, S_2; \mathbf{A}, \mathbf{B})$ (在纯策略意义下) 的一个非合作的平衡点(或纳什平衡点)。

在这里我们就是求对策 $\Gamma = (S_1, S_2; \mathbf{A}, \mathbf{B})$, $S_1 = \{B_i \mid i = 1, 2, \dots, 20\}$, $S_2 = \{G_j \mid j = 1, 2, \dots, 20\}$, $\mathbf{A} = (SS_{ij}^{(1)})_{20 \times 20}$, $\mathbf{B} = (SS_{ji}^{(0)})_{20 \times 20}$ 的纳什平衡点 (i^*, j^*) , 即意味着男青年 B_{i^*} 和女青年 G_{j^*} ($1 \leq i^*, j^* \leq 20$) 配对成功。

按照(15.7)式所确定的算法,分部求平衡点,直到没有纯策略意义下的平衡点为止。编程实现求解,可以得到最佳的配对成功为 10 对,具体如表 15-6。

表 15-6 问题(3)的最优配对方案

| 男 | B_1 | B_3 | B_5 | B_{10} | B_{12} | B_{14} | B_{15} | B_{17} | B_{18} | B_{19} |
|-----|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| 女 | G_{15} | G_{18} | G_{19} | G_9 | G_2 | G_6 | G_4 | G_{14} | G_{11} | G_8 |
| 满意度 | 0.850 7 | 0.909 9 | 0.822 3 | 0.932 7 | 0.887 5 | 0.917 5 | 0.845 9 | 0.897 2 | 0.907 5 | 0.878 4 |

15.6 参考案例与参考文献

1. 参考案例

- (1) 商品价格的确定问题——文献[6]:192—207
- (2) 水雷战问题——文献[7]:215—219
- (3) 作战问题——文献[8]:227—235
- (4) 城市污水的治理问题——文献[9]:235—238
- (5) 体育比赛中的队员出场问题——文献[10]:127—132
- (6) 效益的合理分配问题——文献[11]:179—180

2. 参考文献

- [1] 《运筹学》教材编写组. 运筹学. 修订版. 北京:清华大学出版社,1990
- [2] 杨超等. 运筹学. 北京:科学出版社,2004
- [3] 刘德铭等. 对策论及其应用. 长沙:国防科技大学出版社,1995
- [4] 徐前方等. 对策论. 杭州:浙江大学出版社,2001
- [5] 王建华. 对策论. 北京:清华大学出版社,1986
- [6] 叶其孝. 大学生数学建模竞赛辅导教材(一). 长沙:湖南教育出版

社,1993

- [7] 谭永基等. 数学模型. 上海:复旦大学出版社,1997
- [8] 杨启帆等. 数学建模. 杭州:浙江大学出版社,1999
- [9] 白其峥. 数学建模案例分析. 北京:海洋出版社,2000
- [10] 沈继红等. 数学建模. 哈尔滨:哈尔滨工程大学出版社,1996
- [11] 刘承平. 数学建模方法. 北京:高等教育出版社,2002

第十六章 随机性决策分析方法

人们在日常生活和工作中经常会遇到一些与随机因素有关、后果不确定,而又必须做出判断和决定的问题.这类问题称为**随机性决策问题**.任何一个随机性决策问题都包含两个方面的内容,即决策人所采取的行动方案(简称**决策**)和问题的自然状态(简称**状态**),而且具有两个基本特点:**后果的不确定性和后果的效用**.

所谓后果的不确定性,主要是由于问题的随机性,使得问题会出现什么状态是不确定的,所以决策人做出某种决策以后会出现何种后果也是不确定的.而效用是后果价值的量化,由于不确定性,无论决策人采用什么决策,都可能会遇到事先不能完全预料的后果,这要承担一定的风险,不同的决策人对待冒风险的态度会不同.因而,同样的后果对不同的决策人产生的效用也会不同.即使在没有风险的情况下,不同的决策人对待各种后果也有不同的偏好,为此,在进行定量分析之前,就应该确定出所有后果的效用.只有这样,人们才能比较各种决策的优劣,根据自己的喜好来选择最佳的决策方案.

在决策分析中,后果的不确定性和对于后果赋予的效用是两个关键性的问题.为此,对于状态的不确定性主要用主观概率来表示,而后果的效用则用效用理论来研究.

16.1 随机性决策问题的基本概念

16.1.1 主观概率

随机性决策问题的后果的不确定性,主要是由状态的不确定性所引起的.状态的不确定性,往往不能通过在相同条件下的大量重复试验来确定其概率分布,只能由决策人主观地做出估计,因此称为**主观概率**,它与实际的概率(此称**客观概率**)是有区别的.

主观概率是决策人进行决策分析的依据,虽然它与客观概率有本质的区别,但在定义概率方面有不同之处,同样遵循客观概率应该遵循的若干假设、公理和性质等,因此,适用于客观概率的所有的逻辑推理方法均适用于主观概率.这里仅给出主观概率所服从的基本假设(或称公理系统):

(1) 设 Ω 为一非空集合, 其元素可以是某种试验或观察的结果, 也可以是自然的状态. 将这些元素记作抽象的点 ω , 因而有 $\Omega = \{\omega\}$.

(2) 设 F 是 Ω 中的一些子集 A 所构成的集合, F 满足下列条件:

1) $\Omega \in F$;

2) 如果 $A \in F$, 则 $\bar{A} = \Omega \setminus A \in F$;

3) 如果可列多个 $A_n \in F, n=1, 2, \dots$, 则它们的并集 $\bigcup_{n=1}^{\infty} A_n \in F$.

(3) 设 $P(A) (A \in F)$ 是定义在 F 上的实值集函数, 如果它满足下列条件, 就称为 F 上的 (主观或客观) 概率测度, 或简称概率, 这些条件是

1) 对于每个 $A \in F$, 有 $0 \leq P(A) \leq 1$;

2) $P(\Omega) = 1$;

3) 如果可列多个 $A_n \in F (n=1, 2, \dots)$, $A_i \cap A_j = \emptyset (i \neq j)$, 则

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

这里称点 ω 为基本事件, F 中的集 A 称为事件, F 是全体事件的集合, $P(A)$ 称为事件 A 的 (主观或客观) 概率, 三元总体 (Ω, F, P) 称为 (主观或客观) 概率空间.

设定主观概率的方法主要有: 主观先验分布法、无信息先验分布法、极大熵 (极大平均信息量) 先验分布法和利用过去数据设定先验分布法等^[3,4].

16.1.2 效用函数

在随机性决策问题中, 后果的不确定性是由状态的不确定性引起的. 所以, 在研究后果的效用时要充分考虑后果的不确定性.

设决策人在选择某一行动时, 决策问题可能的 n 个后果为 C_1, C_2, \dots, C_n ; 后果 C_i 可能发生的概率分别为 $p_i (i=1, 2, \dots, n)$, 且 $\sum_{i=1}^n p_i = 1$. 用 P 表示所有后果的概率分布, 并记 $P = (p_1, C_1; p_2, C_2; \dots; p_n, C_n)$, 则称 P 为展望. 所有展望构成的集合记为 \bar{P} , 可以验证 \bar{P} 关于凸线性组合是封闭的, 即如果 $P_1, P_2 \in \bar{P}$, 而且 $0 \leq \lambda \leq 1$, 则有

$$\lambda P_1 + (1 - \lambda) P_2 \in \bar{P}.$$

对于任意两个展望 $P_1, P_2 \in \bar{P}$, 都存在一定的优先关系, 即对于决策人可以认为 P_1 优于 P_2 , 或 P_1 与 P_2 无差异, 或 P_1 不优于 P_2 三种情况, 将这三种关系分别记为 $P_1 > P_2$, $P_1 \sim P_2$ 和 $P_2 \geq P_1$. 这种优先关系反映了决策人对各种后果的偏好程度.

定义 16.1 设 $u(P)$ 是定义在展望 \bar{P} 上的实值函数, 且满足

(1) 它和在 \bar{P} 上的优先关系 \succ 一致, 即如果对于所有 $P_1, P_2 \in \bar{P}$, 有 $P_1 \succ P_2$, 当且仅当 $u(P_1) \geq u(P_2)$;

(2) 它在 \bar{P} 上是线性的, 即如果 $P_1, P_2 \in \bar{P}$, 而且 $0 \leq \lambda \leq 1$, 则

$$u(\lambda P_1 + (1 - \lambda)P_2) = \lambda u(P_1) + (1 - \lambda)u(P_2),$$

那么称 $u(P)$ 是定义在展望 \bar{P} 上的效用函数.

如果 $P = (p_1, C_1; p_2, C_2; \cdots; p_n, C_n) \in \bar{P}$, 则 $u(P)$ 就是表示以概率 p_i 选择 C_i ($i = 1, 2, \cdots, n$) 的期望效用. 效用是决策人在有风险的情况下对后果的偏好的量化, 因此, 其中包含有决策人对于一个不确定事件可能冒风险的态度, 又称这种效用为**基数效用**. 如果所研究的事件是确定的事件, 并不受自然状态的影响, 类似地可以定义一个效用来表示决策人对确定事件的各种后果的偏好程度. 对于这类事件, 决策人无需承担风险, 相应的效用与基数效用有所不同, 在此称之为**序数效用**.

定义 16.2 设 X 为所有确定事件的后果 x 的集合, $u(x)$ 是定义在 X 上的实值函数, 如果对于任意的 $x_1, x_2 \in X$ 有 $u(x_1) \geq u(x_2)$, 当且仅当 $x_1 \succsim x_2$, 则称 $u(x)$ 是定义在 X 上的**序数效用函数**.

基数效用和序数效用的主要区别是: 基数效用在正线性变换下是惟一的, 而序数效用在保序变换下是惟一的.

正线性变换: $\hat{u}(P) = \alpha u(P) + \beta (\alpha > 0)$.

保序变换: $\hat{u}(x) = f(u(x))$, 对任意 $x \in X$, f 为严格单调增加函数.

16.2 效用函数理论

16.2.1 效用与风险的关系

实际中很多的决策问题都涉及经济效益, 对于这类问题, 在后果不确定的情况下, 决策人的决策往往是效益和风险并存. 但对不同的决策人对待风险的态度一般是不同的, 通常可分为三种态度, 即厌恶型、中立型和喜好型.

假设决策人面对一种风险的情况有 $1/2$ 的机会得不到任何盈利, 也有 $1/2$ 的机会盈利 $2a$ 元, 即他的期望盈利为 a 元. 如果决策人认为冒此风险的期望盈利只等价于比它低的不冒风险的盈利, 则对待风险的态度为**厌恶型**的. 否则对待风险的态度为**喜好型**的. 如果决策人认为这和不冒任何风险的另一行为盈利 a 元等价, 则对待风险的态度是**中立型**的. 这三种不同的态度可以反映在效用函数上就是凹(上凸)函数, 线性函数和凸(下凸)函数. 如图 16-1.

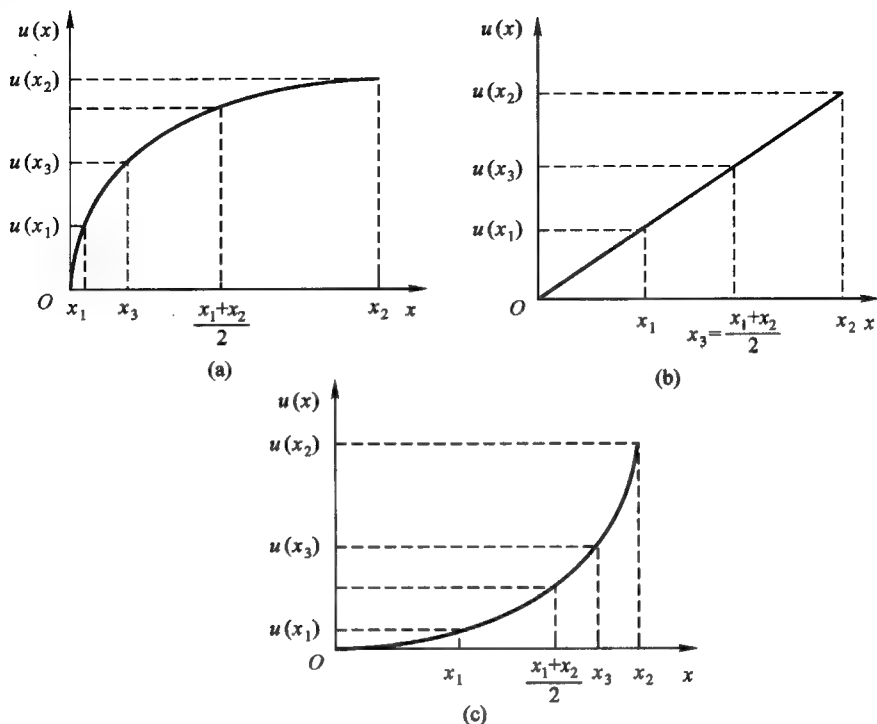


图 16-1 三种不同的效用函数曲线

由图 16-1(a)是风险厌恶型的效用函数,即有

$$u(x_3) = \frac{1}{2}[u(x_1) + u(x_2)] < u\left(\frac{x_1 + x_2}{2}\right);$$

由图 16-1(b)是风险中立型的效用函数,即有

$$u(x_3) = \frac{1}{2}[u(x_1) + u(x_2)] = u\left(\frac{x_1 + x_2}{2}\right);$$

由图 16-1(c)是风险喜好型的效用函数,即有

$$u(x_3) = \frac{1}{2}[u(x_1) + u(x_2)] > u\left(\frac{x_1 + x_2}{2}\right).$$

实际中,很多的情况效用函数的曲线呈 S 型,即在后果的范围内,决策人对待风险的态度往往会从厌恶风险改变为喜好风险.如图 16-2.

图 16-2(a)反映了决策人的财产从小到大,对待风险的态度从喜好到厌恶的改变.图 16-2(b)反映了决策人的财产随着从损失到盈利的增加,对待风险的态度会从喜好到厌恶的变化.这是最常用的效用函数.

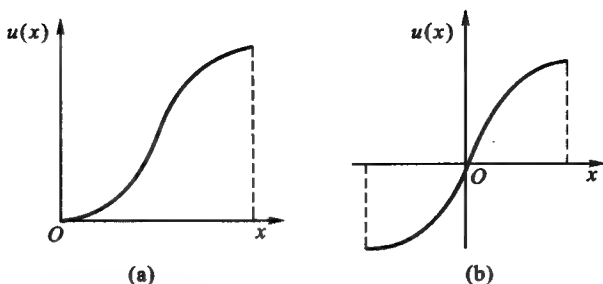


图 16-2 两类 S 型效用函数曲线

16.2.2 损失函数与风险函数

有的时候不用效用函数,而是用损失函数来做决策分析.记损失函数为 $l(x, a)$,它表示一个决策问题当状态为 x ,决策人的行动为 a 时所产生的后果使决策人所受的损失.损失函数可以为正,也可以为负,它反映决策人获得的利益,后果效用愈大,则损失愈小.由此可以用效用函数来定义损失函数,即令

$$l(x, a) = -u(x, a).$$

实际中,在有些问题上为了使损失函数总是非负的,也可以定义损失函数为

$$l(x, a) = \sup_{x \in X} \sup_{a \in A} u(x, a) - u(x, a).$$

在效用理论中,我们说明了期望效用能够合理地表示在风险情况下决策人的偏好,因此,期望损失也必然是决策人在风险情况下遭受损失的一个正确测度.

16.2.3 随机优势与效用函数

随机决策分析是在一定的条件下,用期望效用来表示一个随机事件效用的一种方法.在有关证券问题的研究中,又提出另外一种在一定的风险情况下制订决策的方法,称为随机优势法.

假设问题的效用函数为 $u(x)$,其自变量 x 表示财富(为一随机变量).实际中的问题总是有 $x \in [a, b]$,且 $u(x)$ 在 $[a, b]$ 上有界,对于这种效用函数可以分为以下几类:

1. 递增的效用函数

实际中,一般要求财富的效用函数 $u(x)$ 是 $x \in [a, b]$ 的非递减函数,即意味着当财富增加时,它的效用总不会减少.通常是随着 x 的增加 $u(x)$ 是严格递增的,而且是有界的.为此,我们假设:

- (1) 对于任意 $x_1, x_2 \in [a, b]$, 当 $x_1 < x_2$ 时有 $u(x_1) < u(x_2)$;
- (2) $u(x)$ 在 $[a, b]$ 上连续,且有界,即存在 $M > 0$ 使 $u(x) \leq M$;
- (3) $u(x)$ 在 $[a, b]$ 上一次可微,且在 (a, b) 内有 $0 < u'(x) \leq M$.

记此类效用函数为 U_1 , 即

$$U_1 = \{u \mid u \text{ 和 } u' \text{ 在 } [a, b] \text{ 上连续有界, 且在 } (a, b) \text{ 内 } u' > 0\}$$

这种类型的效用函数仅能反映出财富与风险的关系, 但不能反映出决策人对待风险的态度. 因此 U_1 中既可包含厌恶的效用函数, 也可包含喜好风险和风险中立的效用函数. 为此, 还可以进一步分类.

2. 递增的凹效用函数

这种效用函数是递增的, 故设 $u(x) \in U_1$, 而且是严格凹的, 即 $u(x)$ 在 $[a, b]$ 上具有二阶连续有界的导数. 记为

$$U_2 = \{u \mid u \in U_1, u'' \in C[a, b], \text{ 且在 } (a, b) \text{ 内 } u'' < 0\}.$$

实际中常用的 U_2 类函数有

幂函数: $u(x) = -x^{-c}, x \in [a, b] \ (c > 0, a > 0)$;

对数函数: $u(x) = \ln x, x \in [a, b] \subset (0, +\infty)$;

指数函数: $u(x) = -e^{-cx}, x \in [a, +\infty) \ (c > 0)$.

根据风险和效用的关系, 当 u', u'' 存在, 且 $u' \neq 0$ 时, 定义对待风险态度的局部测度为

$$r(x) = -\frac{u''(x)}{u'(x)},$$

即 $r(x)$ 是效用函数 $u(x)$ 的曲率测度. 可以证明: 如果 $r(x) > 0$, 则决策人的财产为 x 时, 他是厌恶风险的. 如果 $r(x) = 0$, 则决策人的财产为 x 时, 他是风险中立的. 如果 $r(x) < 0$, 则决策人的财产为 x 时, 他是追求风险的, 而且 $|r(x)|$ 愈大, 他愈厌恶 (或追求) 风险.

3. 递减的厌恶风险的效用函数

实际中, 多数决策人对小额盈亏的态度是随着财富的积累而变化的, 他们的财富积累愈多, 对小额盈亏所冒风险的厌恶程度愈小. 因此, 我们假设 $r(x)$ 是 x 的非递增的函数, 则可以得到一类效用函数, 记为

$$U_3 = \{u \mid u \in U_2, r(x) \text{ 在 } [a, b] \text{ 上连续可微, 有界, 且 } r'(x) \leq 0\},$$

即 U_3 是 U_2 的一个子类.

由于当 $r'(x) \leq 0$ 时, $r(x)$ 是非递增的. 要使 $r'(x) \leq 0$, 即

$$r'(x) = -\frac{u'''u' - [u''(x)]^2}{[u'(x)]^2} \leq 0,$$

则 $u'''u' - (u'')^2 \geq 0$, 故 $u'''(x) - \frac{(u'')^2}{u'} \geq 0$. 因此, U_3 类函数存在的必要条件是 $u''' > 0, x \in (a, b)$, 但不是充分条件.

上面给出了适应于不同情况的效用函数的基本形式, 实际中需要依据具体问题的性质, 来选用合适的效用函数, 对问题进行研究.

16.3 彩票中的数学问题^{①[7]}

16.3.1 问题的提出

目前流行的彩票主要有“传统型”和“乐透型”两种类型.

“传统型”采用“10 选 6 + 1”方案:先从 6 组 0 ~ 9 号球中摇出 6 个基本号码,每组摇出一个,然后从 0 ~ 4 号球中摇出一个特别号码.投注者从 0 ~ 9 十个号码中任选 6 个基本号码(可重复),从 0 ~ 4 中选一个特别号码,构成一注,根据单注号码与中奖号码相符的个数多少及顺序确定中奖等级.以中奖号码“abcdef + g”为例说明中奖等级,如表 16 - 1.

表 16 - 1 传统型彩票方案的设奖规则

| 中奖等级 | 10 选 6 + 1(6 + 1/10) | | |
|------|------------------------------------|------|---------------|
| | 基 本 号 码 | 特别号码 | 说明 |
| 一等奖 | abcdef | g | 选 7 中 (6 + 1) |
| 二等奖 | abcdef | | 选 7 中 (6) |
| 三等奖 | abcde× ×bcdef | | 选 7 中 (5) |
| 四等奖 | abcd×× ×bcde× ××cdef | | 选 7 中 (4) |
| 五等奖 | abc××× ×bcd×× ××cde× ×××def | | 选 7 中 (3) |
| 六等奖 | ab×××× ×bc××× ××cd×× ×××de× ××××ef | | 选 7 中 (2) |

注:×表示未选中的号码

“乐透型”有多种不同的形式,比如“33 选 7”的方案:先从 01 ~ 33 个号码球中一个一个地摇出 7 个基本号,再从剩余的 26 个号码球中摇出一个特别号码.投注者从 01 ~ 33 个号码中任选 7 个组成一注(不可重复),根据单注号码与中奖号码相符的个数多少确定相应的中奖等级,不考虑号码顺序.又如“36 选 6 + 1”的方案,先从 01 ~ 36 个号码球中摇出 6 个基本号,再从剩下的 30 个号码球中摇出一个特别号码.投注者从 01 ~ 36 个号码中任选 7 个组成一注(不可重复),根据单注号码与中奖号码相符的个数多少确定相应的中奖等级,不考虑号码顺序.这两种方案的中奖等级如表 16 - 2.

① 本案例选自 2002 年中国大学生数学建模竞赛 B 题

表 16-2 乐透型彩票方案的设奖规则

| 中奖等级 | 33 选 7 (7/33) | | | 36 选 6+1 (6+1/36) | | |
|------|---------------|------|-------------|-------------------|------|-------------|
| | 基本号码 | 特别号码 | 说明 | 基本号码 | 特别号码 | 说明 |
| 一等奖 | ●●●●●●● | | 选 7 中 (7) | ●●●●●●★ | | 选 7 中 (6+1) |
| 二等奖 | ●●●●●●○ | ★ | 选 7 中 (6+1) | ●●●●●● | | 选 7 中 (6) |
| 三等奖 | ●●●●●●○ | | 选 7 中 (6) | ●●●●●○★ | | 选 7 中 (5+1) |
| 四等奖 | ●●●●●○● | ★ | 选 7 中 (5+1) | ●●●●●○ | | 选 7 中 (5) |
| 五等奖 | ●●●●●○● | | 选 7 中 (5) | ●●●●○○★ | | 选 7 中 (4+1) |
| 六等奖 | ●●●●○○● | ★ | 选 7 中 (4+1) | ●●●●○○ | | 选 7 中 (4) |
| 七等奖 | ●●●●○○○ | | 选 7 中 (4) | ●●●○○○★ | | 选 7 中 (3+1) |

注：● 为选中的基本号码；★ 为选中的特别号码；○ 为未选中的号码

以上两种类型的奖金总额一般为销售总额的 50%，投注者单注金额为 2 元，单注若已获得高级别的奖就不再兼得低级别的奖。现在常见的销售规则及相应的奖金设置方案如表 16-3，其中一、二、三等奖为高项奖，后面的为低项奖。低项奖数额固定，高项奖按比例分配，但一等奖单注保底金额 60 万元，封顶金额 500 万元，各高项奖额的计算方法为

$$[(\text{当期销售总额} \times \text{总奖金比例}) - \text{低项奖总额}] \times \text{单项奖比例}.$$

表 16-3 现有的彩票方案

| 序号 | 奖项方案 | 一等奖比例 | 二等奖比例 | 三等奖比例 | 四等奖金额 | 五等奖金额 | 六等奖金额 | 七等奖金额 | 备注 |
|----|--------|-------|-------|-------|-------|-------|-------|-------|----|
| 1 | 6+1/10 | 50% | 20% | 30% | 50 | | | | 按序 |
| 2 | 6+1/10 | 60% | 20% | 20% | 300 | 20 | 5 | | 按序 |
| 3 | 6+1/10 | 65% | 15% | 20% | 300 | 20 | 5 | | 按序 |
| 4 | 6+1/10 | 70% | 15% | 15% | 300 | 20 | 5 | | 按序 |
| 5 | 7/29 | 60% | 20% | 20% | 300 | 30 | 5 | | |
| 6 | 6+1/29 | 60% | 25% | 15% | 200 | 20 | 5 | | |
| 7 | 7/30 | 65% | 15% | 20% | 500 | 50 | 15 | 5 | |
| 8 | 7/30 | 70% | 10% | 20% | 200 | 50 | 10 | 5 | |
| 9 | 7/30 | 75% | 10% | 15% | 200 | 30 | 10 | 5 | |
| 10 | 7/31 | 60% | 15% | 25% | 500 | 50 | 20 | 10 | |
| 11 | 7/31 | 75% | 10% | 15% | 320 | 30 | 5 | | |

续表

| 序号 | 奖项方案 | 一等奖比例 | 二等奖比例 | 三等奖比例 | 四等奖金额 | 五等奖金额 | 六等奖金额 | 七等奖金额 | 备注 |
|----|----------|-------|-------|-------|-------|-------|-------|-------|-----|
| 12 | 7/32 | 65% | 15% | 20% | 500 | 50 | 10 | | |
| 13 | 7/32 | 70% | 10% | 20% | 500 | 50 | 10 | | |
| 14 | 7/32 | 75% | 10% | 15% | 500 | 50 | 10 | | |
| 15 | 7/33 | 70% | 10% | 20% | 600 | 60 | 6 | | |
| 16 | 7/33 | 75% | 10% | 15% | 500 | 50 | 10 | 5 | |
| 17 | 7/34 | 65% | 15% | 20% | 500 | 30 | 6 | | |
| 18 | 7/34 | 68% | 12% | 20% | 500 | 50 | 10 | 2 | |
| 19 | 7/35 | 70% | 15% | 15% | 300 | 50 | 5 | | |
| 20 | 7/35 | 70% | 10% | 20% | 500 | 100 | 30 | 5 | |
| 21 | 7/35 | 75% | 10% | 15% | 1 000 | 100 | 50 | 5 | |
| 22 | 7/35 | 80% | 10% | 10% | 200 | 50 | 20 | 5 | |
| 23 | 7/35 | 100% | 2 000 | 20 | 4 | 2 | | | 无特号 |
| 24 | 6 + 1/36 | 75% | 10% | 15% | 500 | 100 | 10 | 5 | |
| 25 | 6 + 1/36 | 80% | 10% | 10% | 500 | 100 | 10 | | |
| 26 | 7/36 | 70% | 10% | 20% | 500 | 50 | 10 | 5 | |
| 27 | 7/37 | 70% | 15% | 15% | 1 500 | 100 | 50 | | |
| 28 | 6/40 | 82% | 10% | 8% | 200 | 10 | 1 | | |
| 29 | 5/60 | 60% | 20% | 20% | 300 | 30 | | | |

要解决的问题是

- (1) 根据这些方案的具体情况,试研究各种奖项出现的可能性;
- (2) 根据问题中所给方案的具体情况,综合分析各种奖项出现的可能性、奖项和奖金额的设置以及对彩民的吸引力等因素评价各方案的合理性;
- (3) 设计一种“更好”的方案及相应的算法,并据此给彩票管理部门提出建议.

16.3.2 模型的假设与符号说明

1. 模型的假设

- (1) 彩票摇奖是公平公正的,各号码的出现是随机的;
- (2) 彩民购买彩票是随机的独立事件;

(3) 对同一方案中高级别奖项的奖金比例或奖金额不应低于相对低级别的奖金比例或奖金额;

(4) 根据我国的现行制度,假设我国居民的平均工作年限为 $T=35$ 年.

2. 符号说明

r_j 表示第 j ($j=1,2,3$) 等(高项)奖占高项奖总额的比例; x_i 表示第 i ($1 \leq i \leq 7$) 等奖奖金额的均值; p_i 表示彩民中第 i 等奖 x_i ($1 \leq i \leq 7$) 的概率; $\mu(x_i)$ 表示彩民对某个方案第 i 等奖的满意度,即第 i ($1 \leq i \leq 7$) 等奖对彩民的吸引力; λ 表示某地区的平均收入和消费水平的相关因子,称为“实力因子”,一般为常数; F 表示彩票方案的合理性指标,即方案设置对彩民吸引力的综合指标.

16.3.3 问题的分析

从目前全国各省(市、自治区)的彩票发行情况来看,其规则、奖项设置和设奖比例都不尽相同,而且有的差异很大,运行模式不统一,管理还不够规范.为此,我们认为对中国目前的彩票市场的运作情况进行研究和评价是必要的,尤其是对目前已有的彩票规则的合理性评估,以及现行规则是否符合本地区的实际情况,能否通过彩票发行规则的制定提高对广大彩民的吸引力,促使更多的人加入到彩民的行列中来,使得国家和彩民的利益得到双赢,进一步促进我国的彩票事业的健康发展.

评价一个方案的优劣,或合理性如何,主要取决于彩票公司和广大彩民两方面的利益.事实上,公司和彩民各得销售总额的 50% 是确定的,双方的利益主要就取决于销售总额的大小,即双方的利益都与销售额成正比.因此,问题是怎样才能有利于销售额的增加?即公司采用什么样的方案才能吸引广大的彩民积极踊跃购买彩票?具体地讲,问题涉及一个方案的设置使彩民获奖的可能性有多大、奖金额有多少、中奖面怎样、各奖项的设置是否合理等因素,这些因素都对彩民购买彩票的吸引力产生一定的影响,在这里可用彩民的心理曲线来描述一个方案对彩民的吸引力变化.另外,一个方案对彩民的影响程度可能与区域有关,即与彩民所在地区的经济状况以及收入和消费水平有关.为此,我们要考查一个方案的合理性问题,需要综合考虑以上这些因素的影响,这是我们建立模型的关键所在.

16.3.4 模型的准备

1. 确定彩民的心理曲线

一般说来,人们对待问题的心理变化过程是非常复杂,而且也是模糊的.在此,彩民对一个方案的各个奖项及奖金额的看法(即对彩民的吸引力)的变化就是一个典型的模糊概念.由模糊数学隶属度的概念和心理学的相关知识,

根据人们通常对一件事物的心理变化一般遵循的规律,不妨定义彩民的心理曲线为

$$\mu(x) = 1 - e^{-\left(\frac{x}{\lambda}\right)^2} \quad (\lambda > 0),$$

其中 λ 表示彩民平均收入的相关因子,称为实力因子,一般为常数.

2. 计算实力因子 λ

实力因子是反应一个地区的彩民的平均收入和消费水平的指标,确定一个地区的彩票方案应该考虑所在地区的实力因子,在我国不同地区的收入和消费水平是不同的,因此,不同地区的实力因子应有一定的差异,目前各地区现行的方案不尽相同,要统一来评估这些方案的合理性,就应该对同一个实力因子进行研究.为此,我们以中等地区的收入水平(或全国平均水平)为例进行研究.根据相关网站的统计数据,不妨取人均年收入 1.5 万元,按我国的现行制度,平均工作年限 $T = 35$ 年,则人均总收入为 52.5 万元,于是,当 $x_0 = 52.5$ 万元时,取 $\mu(x_0) = 1 - e^{-\left(\frac{x_0}{\lambda}\right)^2} = 0.5$ (即吸引力的中位数),则有

$$\lambda = \frac{5.25 \times 10^5}{\sqrt{-\ln 0.5}} \approx 6.30589 \times 10^5.$$

同理,可以算出年收入 1 万元、2 万元、2.5 万元、3 万元、4 万元、5 万元、10 万元的实力因子如表 16-4.

表 16-4 不同情况的实力因子

| 年 收入 指 标 | 1 万元 | 1.5 万元 | 2 万元 | 2.5 万元 | 3 万元 | 4 万元 | 5 万元 | 10 万元 |
|-------------------|---------|---------|---------|-----------|-----------|-----------|-----------|-----------|
| λ | 420 393 | 630 589 | 840 786 | 1 050 982 | 1 261 179 | 1 681 571 | 2 101 964 | 4 203 928 |

16.3.5 模型的建立与求解

问题(1): 研究所有方案各种奖项出现的可能性.

从已给的 29 种方案可知,可将其分为四类. K_1 : 10 选 6 + 1 (6 + 1/10) 型; K_2 : n 选 m (m/n) 型; K_3 : n 选 $m + 1$ ($m + 1/n$) 型和 K_4 : n 选 m (m/n) 无特别号型,下面分别给出各种类型方案出现各奖项的概率计算公式.

(1) K_1 : 10 选 6 + 1 (6 + 1/10) 型

$$p_1 = \frac{1}{5 \times 10^6} = 2 \times 10^{-7}, \quad p_2 = \frac{4}{5 \times 10^6} = 8 \times 10^{-7},$$

$$p_3 = \frac{2 \times C_9^1}{10^6} = \frac{2 \times 9}{10^6} = 1.8 \times 10^{-5},$$

$$p_4 = \frac{2C_9^1 C_{10}^1 + C_9^1 C_9^1}{10^6} = \frac{2 \times 9 \times 10 + 9 \times 9}{10^6} = 2.61 \times 10^{-4},$$

$$p_5 = \frac{2C_9^1 C_{10}^1 C_{10}^1 + 2C_9^1 C_9^1 C_{10}^1}{10^6} = \frac{2 \times 9 \times 10^2 + 2 \times 9^2 \times 10}{10^6} = 3.42 \times 10^{-3},$$

$$p_6 = \frac{2 \times C_9^1 C_{10}^1 C_{10}^1 C_{10}^1 + 3 \times C_9^1 C_9^1 C_{10}^1 C_{10}^1 - (3 \times C_9^1 \times C_9^1 + 2 \times C_9^1)}{10^6}$$

$$= \frac{2 \times 9 \times 10^3 + 3 \times 9^2 \times 10^2 - (3 \times 9^2 + 2 \times 9)}{10^6} = 4.2039 \times 10^{-2}.$$

(2) $K_2: n$ 选 $m(m/n)$ 型

$$p_1 = \frac{1}{C_n^m}, \quad p_2 = \frac{C_m^{m-1}}{C_n^m}, \quad p_3 = \frac{C_m^{m-1} C_{n-(m+1)}^1}{C_n^m}, \quad p_4 = \frac{C_m^{m-2} C_{n-(m+1)}^1}{C_n^m},$$

$$p_5 = \frac{C_m^{m-2} C_{n-(m+1)}^2}{C_n^m}, \quad p_6 = \frac{C_m^{m-3} C_{n-(m+1)}^2}{C_n^m}, \quad p_7 = \frac{C_m^{m-3} C_{n-(m+1)}^3}{C_n^m}.$$

(3) $K_3: n$ 选 $m+1(m+1/n)$ 型

$$p_1 = \frac{1}{C_n^{m+1}}, \quad p_2 = \frac{C_m^1 C_{n-(m+1)}^1}{C_n^{m+1}}, \quad p_3 = \frac{C_m^{m-1} C_{n-(m+1)}^1}{C_n^{m+1}}, \quad p_4 = \frac{C_m^{m-1} C_{n-(m+1)}^2}{C_n^{m+1}},$$

$$p_5 = \frac{C_m^{m-2} C_{n-(m+1)}^2}{C_n^{m+1}}, \quad p_6 = \frac{C_m^{m-2} C_{n-(m+1)}^3}{C_n^{m+1}}, \quad p_7 = \frac{C_m^{m-3} C_{n-(m+1)}^3}{C_n^{m+1}}.$$

(4) $K_4: n$ 选 $m(m/n)$ 无特别号型

$$p_1 = \frac{1}{C_n^m}, \quad p_2 = \frac{C_m^{m-1} C_{n-m}^1}{C_n^m}, \quad p_3 = \frac{C_m^{m-2} C_{n-m}^2}{C_n^m},$$

$$p_4 = \frac{C_m^{m-3} C_{n-m}^3}{C_n^m}, \quad p_5 = \frac{C_m^{m-4} C_{n-m}^4}{C_n^m}.$$

各种方案的各个奖项获奖概率及获奖总概率 $P = \sum_i p_i$, 计算结果如表

16-5. 由此可确定各方案的各个奖项出现的可能性.

表 16-5 各方案的设奖概率

| 序号 | 概 率 方 案 | $p_1(10^{-7})$ | $p_2(10^{-6})$ | $p_3(10^{-5})$ | $p_4(10^{-4})$ | $p_5(10^{-3})$ | $p_6(10^{-2})$ | p_7 | $p = \sum_i p_i$ |
|-------|---------------|--------------------|----------------|----------------|----------------|----------------|----------------|-----------|------------------|
| 1-4 | 6+1/10 | 2×10^{-7} | 0.8 | 1.8 | 2.61 | 3.42 | 4.230 9 | ... | 0.045 695 |
| 5 | 7/29 | 6.407 05 | 4.484 94 | 9.418 4 | 2.825 5 | 2.825 5 | 0.470 92 | 0.029 825 | 0.037 742 |
| 6 | 6+1/29 | 6.407 05 | 1.409 6 | 8.457 3 | 8.880 2 | 2.220 0 | 1.480 0 | 0.019 734 | 0.037 742 |
| 7-9 | 7/30 | 4.912 07 | 3.438 45 | 7.564 6 | 2.269 4 | 2.382 8 | 0.397 14 | 0.026 476 | 0.033 137 |
| 10-11 | 7/31 | 3.802 90 | 2.662 03 | 6.122 7 | 1.836 8 | 2.020 5 | 0.336 75 | 0.023 572 | 0.029 208 |
| 12-14 | 7/32 | 2.971 01 | 2.079 71 | 4.991 3 | 1.497 4 | 1.722 | 0.287 00 | 0.021 047 | 0.025 832 |
| 15-16 | 7/33 | 2.340 80 | 1.638 56 | 4.096 4 | 1.228 9 | 1.474 7 | 0.245 78 | 0.018 843 | 0.022 941 |
| 17-18 | 7/34 | 1.858 87 | 1.301 21 | 3.383 1 | 1.014 9 | 1.268 7 | 0.211 45 | 0.016 916 | 0.020 436 |
| 19-22 | 7/35 | 1.487 09 | 1.040 97 | 2.810 6 | 0.843 18 | 1.096 1 | 0.182 69 | 0.015 224 | 0.018 261 |
| 23 | 7/35 | 1.487 09 | 29.147 | 118.05 | 170.51 | 106.57 | ... | ... | 0.124 83 |
| 24-25 | 6+1/36 | 1.197 94 | 3.474 02 | 2.084 4 | 2.918 2 | 0.729 54 | 0.656 59 | 0.008 755 | 0.016 367 |
| 26 | 7/36 | 1.197 94 | 0.838 556 | 2.348 0 | 0.704 39 | 0.950 92 | 0.158 49 | 0.013 736 | 0.016 367 |
| 27 | 7/37 | 0.971 30 | 0.679 911 | 1.971 7 | 0.591 52 | 0.828 13 | 0.138 02 | 0.012 422 | 0.014 710 |
| 28 | 6/40 | 2.605 3 | 1.563 2 | 5.158 4 | 1.289 6 | 2.063 4 | 0.275 12 | 0.028 428 | 0.033 425 |
| 29 | 5/60 | 1.831 | 0.915 5 | 4.943 7 | 0.988 74 | 2.620 2 | 0.262 02 | 0.045 416 | 0.050 806 |

问题(2): 综合分析评价各方案的合理性.

要综合评价这些已知方案的合理性, 应该建立一个能够充分反应各种因素影响的合理性指标函数. 因为彩民购买彩票可以认为是一种冒险行为, 为此, 我们根据随机决策分析中风险决策的理论, 考虑到彩民的心理因素的影响, 可取

$$\mu(x) = 1 - e^{-\left(\frac{x}{\lambda}\right)^2} \quad (\lambda > 0)$$

为随机性决策的效用函数, 如图 16-3 所示. 于是取如下的指标函数

$$F = \sum_{i=1}^7 p_i \mu(x_i), \quad (16.1)$$

即表示在考虑彩民的心理因素的条件下, 一个方案的中奖率、中奖面、奖项和奖金设置等因素对彩民的吸引力.

另一方面, 由题意知, 单注所有可能的低项奖金总额为 $L = \sum_{i=4}^7 p_i x_i$, 根据高项奖的计算公式得单注可能的第 j 项奖(高项奖)金额为

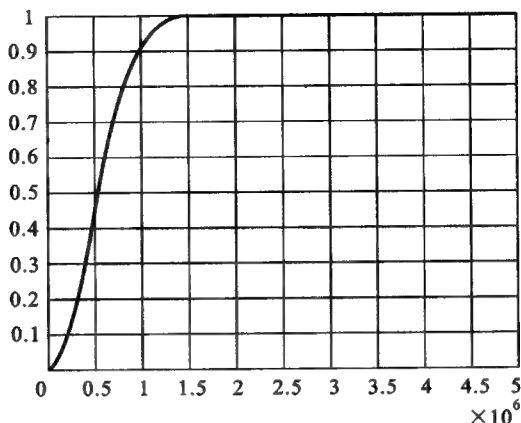


图 16-3 效用函数曲线

$$p_jx_j = (1 - L)r_j = \left(1 - \sum_{i=4}^7 p_ix_i\right)r_j, \quad j = 1, 2, 3,$$

故平均值为

$$x_j = \frac{\left[(1 - \sum_{i=4}^7 p_ix_i)r_j\right]}{p_j}, \quad j = 1, 2, 3. \tag{16.2}$$

于是由(16.1),(16.2)式得

$$\begin{cases} F = \sum_{i=1}^7 p_i\mu(x_i), \\ x_j = \frac{\left[(1 - \sum_{i=4}^7 p_ix_i)r_j\right]}{p_j}, \quad j = 1, 2, 3, \\ \mu(x_i) = 1 - e^{-\left(\frac{x_i}{\lambda}\right)^2}, \quad i = 1, 2, \dots, 7, \\ \lambda = 6.305\,89 \times 10^5. \end{cases} \tag{16.3}$$

利用 MATLAB 可计算出 29 种方案的合理性指标值 F 及高项奖的期望值, 排在前三位的如下表 16-6, 其他各方案也都有相应的结果.

表 16-6 较合理的三种方案

| 指 标 | | F | x_1 | x_2 | x_3 | 排 序 |
|-----|------|------------------------|---------------------|--------|-------|-----|
| 方 案 | | | | | | |
| 9 | 7/30 | 4.009×10^{-7} | 1.086×10^6 | 20 679 | 1 410 | 1 |
| 11 | 7/31 | 3.784×10^{-7} | 1.704×10^6 | 32 448 | 2 116 | 2 |
| 5 | 7/29 | 3.637×10^{-7} | 7.557×10^5 | 35 984 | 1 714 | 3 |

问题(3): 设计一种“更好”的方案.

根据问题(2)的讨论, 现在的问题是取什么样的方案 m/n (n 和 m 取何值)、设置哪些奖项、高项奖的比例 r_j ($j=1, 2, 3$) 为多少和低项奖的奖金额 x_i ($i=4, 5, 6, 7$) 为多少时, 使目标函数 $F = \sum_{i=1}^7 p_i \mu(x_i)$ 有最大值.

设以 m, n, r_j ($j=1, 2, 3$), x_i ($i=4, 5, 6, 7$) 为决策变量, 以它们之间所满足的关系为约束条件, 则可得到非线性规划模型:

$$\begin{aligned} \max F &= \sum_{i=1}^7 p_i \mu(x_i), \\ \text{s. t. } \begin{cases} x_j = \frac{[(1 - \sum_{i=4}^7 p_i x_i) r_j]}{p_j}, j = 1, 2, 3, & (1) \\ \mu(x_i) = 1 - e^{-\left(\frac{x_i}{\lambda}\right)^2} (i = 1, 2, \dots, 7), \lambda = 6.30589 \times 10^5, & (2) \\ r_1 + r_2 + r_3 = 1, & (3) \\ 0.5 \leq r_1 \leq 0.8, & (4) \\ 6 \times 10^5 \leq x_1 \leq 5 \times 10^6, & (5) \\ a_i \leq \frac{x_i}{x_{i+1}} \leq b_i, i = 1, 2, \dots, 6, & (6) \\ p_i < p_{i+1}, i = 1, 2, \dots, 6, & (7) \\ 5 \leq m \leq 7, & (8) \\ 29 \leq n \leq 60, & (9) \\ r_j > 0, x_i \geq 0; m, n \text{ 为正整数.} \end{cases} \end{aligned}$$

关于约束条件的说明:

(1) 条件(1), (2)同问题(2);

(2) 条件(3), (4)是对高项奖的比例约束, r_1 的值不能太大或太小, (4)是根据已知的方案确定的;

(3) 条件(5)是根据题意中一等奖的保底额和封顶额确定的;

(4) 条件(6)中的 a_i, b_i ($i=1, 2, \dots, 6$) 分别为 i 等奖的奖金额 x_i 比 $i+1$ 等奖的奖金额 x_{i+1} 高的倍数, 可由问题(2)的计算结果和已知各方案的奖金数额统计得

$$\begin{aligned} a_1 &= 10, b_1 = 233; a_2 = 4, b_2 = 54; a_3 = 3; b_3 = 17; a_4 = 4, \\ b_4 &= 20; a_5 = 2, b_5 = 10; a_6 = 2, b_6 = 10; \end{aligned}$$

(5) 条件(7)是根据实际问题确定的, 实际中高等奖的概率 p_i 应小于低等奖

的概率 p_{i+1} ,它的值主要由 m, n 确定;

(6) 条件(8),(9)是对方案中 m, n 取值范围的约束,是由已知的方案确定的.

这是一个较复杂的非线性(整数)规划,其中概率 p_i 的取值分为四种不同的情况 K_1, K_2, K_3, K_4 ,都由整数变量 m, n 确定.这个模型的求解是困难的.为此,利用 MATLAB 可求得最优解为 $\{K_2, 6, 32, 0.8, 0.11, 0.09, 200, 10, 1, 0\}$,最优值为 $F = 6.8399 \times 10^{-7}$.故对应的最优方案为32选6(6/32),一、二、三等奖的比例分别为80%、11%、9%、四、五、六、七等奖的金额分别为200、10、1、0元.

前面是针对中等收入水平的彩民情况考虑的,对于经济发达地区和欠发达地区应有所不同.这里分别对年收入1万元、2万元、2.5万元、3万元、4万元、5万元、10万元,工作年限均35年的情况进行了讨论,给出适用于相应各种情况的最优方案,如表16-7.

表 16-7 不同地区的最优方案

| 年 收入 指 标 | 1 万元 | 2 万元 | 2.5 万元 | 3 万元 | 4 万元 | 5 万元 | 10 万元 |
|----------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| λ | 420 393 | 840 786 | 1 050 982 | 1 261 179 | 1 681 571 | 2 101 964 | 4 203 928 |
| 最优方案 | 5 + 1/33 | 6/32 | 7/30 | 6/37 | 6 + 1/32 | 7/33 | 7/35 |
| F | 8.255×10^{-7} | 4.623×10^{-7} | 4.103×10^{-7} | 3.223×10^{-7} | 2.475×10^{-7} | 2.075×10^{-7} | 1.828×10^{-7} |
| r_1 | 0.80 | 0.80 | 0.73 | 0.70 | 0.73 | 0.73 | 0.80 |
| r_2 | 0.10 | 0.11 | 0.17 | 0.15 | 0.19 | 0.18 | 0.13 |
| r_3 | 0.10 | 0.9 | 0.10 | 0.15 | 0.07 | 0.09 | 0.07 |
| x_1 | 6.5×10^5 | 6.18×10^5 | 1.38×10^6 | 1.46×10^6 | 2.23×10^6 | 2.99×10^6 | 3.91×10^6 |
| x_2 | 3 037 | 120 004 | 47 506 | 52 172 | 22 721 | 1.07×10^5 | 94 252 |
| x_3 | 607 | 600 | 1 235 | 1 739 | 1 507 | 1 974 | 1 746 |
| x_4 | 138 | 200 | 100 | 200 | 100 | 200 | 103 |
| x_5 | 7 | 10 | 10 | 20 | 20 | 10 | 20 |
| x_6 | 1 | 1 | 5 | 2 | 2 | 2 | 5 |
| x_7 | 0 | 0 | 0 | 0 | 0 | 0 | 3 |

16.3.6 其他可行的方法

主要典型的方法可以归纳为三大类:

1. 风险决策的效用函数法

利用风险决策的理论和方法对问题进行研究,但在效用函数的取法和合理

性指标函数的确定上可以有所不同,这里介绍几种.

(1) 取偏大型正态分布函数 $A(x) = 1 - e^{-\left(\frac{x}{\sigma}\right)^2}$ ($x > 0$) 为效用函数,并将彩票方案对彩民的吸引力分为奖金的吸引力和中奖率的吸引力两个部分,分别构造出了两个满意度函数 $A(S)$ 和 $A(p)$,并依据国家权威部门的调查报告中的数据将彩民分为冒险者、中立者和避险者三类,比例分别为 $a_1 = 37\%$, $a_2 = 32\%$, $a_3 = 31\%$,按照各自的最大满意度分别确定了相应的满意参数 σ_i ($i = 1, 2, 3$),于是得到了满意度函数

$$\varphi_i(s) = 1 - e^{-\left(\frac{s}{\sigma_i}\right)^2}, \psi_i(p) = 1 - e^{-\left(\frac{p}{\sigma_i}\right)^2} (i = 1, 2, 3),$$

由此构造出吸引度函数

$$f = a_1[\varphi_1(s_1) + \psi_1(p_1)] + a_2[\varphi_2(\bar{s}_{23}) + \psi_2(p_{(13)})] \\ + a_3[\varphi_3(\bar{s}_{47}) + \psi_3(p_{(17)})],$$

其中 \bar{s}_{ij} 表示第 i 等奖到第 j 等奖的奖金的平均值, $p_{(ij)}$ 表示第 i 等奖到第 j 等奖的概率之和.

以 f 为合理性指标函数对已有方案进行评价,可以解决问题(2). 并对各奖项的设置、奖金比例、奖金数额和相应的概率给出约束,建立了相应的非线性优化模型,可以解决问题(3).

(2) 取博彩的心理函数为 $w(t) = 1 - e^{-nt}$, 其中 $t = E\left(\frac{\xi}{n}\right)$, 即表示单注彩票的平均收益, n 为销售注数. 于是构造吸引力函数为 $f = w(t) \cdot \mu$, 这里 $\mu = 7\sqrt{d_1 d_2}$ 表示总公平因子, $d_1 = \prod_{i=1}^3 p_i x_i$, $d_2 = \prod_{i=4}^7 p_i x_i$, 即高项奖和低项奖的公平因子. 并以 f 为指标函数建立了相应的优化模型.

(3) 取效用函数为 $g(x) = \frac{x}{x+a}$, 参数 a 利用“高通滤波系统中下限截止频率”的方法可以确定为 $a = 0.2$.

(4) 取模糊数学中隶属度函数

$$u(x) = \begin{cases} 1 - e^{-\left(\frac{x}{\sigma}\right)^2}, & x > 0, \\ e^{-\left(\frac{x}{\sigma}\right)^2} - 1, & x < 0 \end{cases}$$

来反映彩票方案对彩民影响的心理变化规律,这里 $x < 0$ 说明当彩民购买彩票没有获奖而对彩民的负面影响的程度. 并分别考虑客观概率 p_i 和主观概率 $v(p_i)$ ($i = 1, 2, \dots, 7$),相应地构造出合理性指标函数

$$U_1 = \sum_{i=1}^7 p_i u(s_i - 2) \text{ 和 } U_2 = \sum_{i=1}^7 v(p_i) u(s_i - 2).$$

所谓的主观概率,主要是从心理学的角度“人们通常倾向于高估低概率事件的出现,低估高概率事件的出现”,这就是主观概率与客观概率的差异.由此进一步构造出了合理性指标函数和优化模型.

2. 层次分析法

层次分析主要是可以对问题所给的各方案进行研究评价,但所考虑的因素多,主观意识强,一般认为这种方法缺少可靠的依据,不是十分有效的方法,使用得当也只能对问题(2)进行讨论,解决问题(3)却无能为力.

3. 分类加权法

将彩民分为风险喜好型、风险厌恶型和中性型三类,不同类型的彩民对各因素的看法不同,一般认为用这种方法也有一定的主观性.

16.3.7 进一步要研究的问题

我国的彩票事业还处于发展的阶段,还有一些复杂的问题需要进一步的研究.就本题而言,也有一些值得进一步研究的问题:

(1) 关于彩票方案对彩民的吸引力问题,即彩民对待某种彩票方案的心理状态如何?用什么样的心理曲线能够准确地反映出不同类型彩民的心理变化情况?

(2) 合理性指标函数的构造问题;

(3) 彩票公司发行彩票的收益和风险问题,以及彩民购买彩票的中奖与风险的关系问题;

(4) 奖池有奖金的滚动积累,对彩票的发行和对彩民的心理影响问题.

16.4 参考案例与参考文献

1. 参考案例

- (1) 工艺品的着色问题——文献[3]:17—20
- (2) 新杀虫剂的研制问题——文献[3]:89—99
- (3) 人在应对各种情况的心理变化问题——文献[3]:100—114
- (4) 投资决策问题——文献[4]:107—121
- (5) 电力系统扩建问题——文献[4]:136—139

2. 参考文献

- [1] 杨超等. 运筹学. 北京:科学出版社,2004
- [2] 何坚勇. 运筹学基础. 北京:清华大学出版社,2000

- [3] 陈 珽. 决策分析. 北京:科学出版社,1997
- [4] 岳超源. 决策理论与方法. 北京:科学出版社,2003
- [5] 叶雅阁等. 决策科学手册. 天津:天津科技翻译出版社,1989
- [6] 么焕民等. 数学建模. 下册. 哈尔滨:哈尔滨工业大学出版社,2003
- [7] 韩中庚. “彩票中的数学”问题的优化模型与评述. 工程数学学报, 2003,20(5):107—116

第十七章 多目标决策分析方法

实际中,许多决策问题都是属于多目标决策问题,特别是在工程系统和社会系统等领域,大量的决策问题都有多个目标,具有两个和两个以上目标的决策问题统称为多目标决策问题。例如,一家公司(或企业)要确定下一年度的投资组合方案问题,在可供选择的多个候选方案中,往往都是收益与风险并存,而且收益越高,风险也就越大。如何选择合适的投资组合方案,使收益最高,风险最小呢?这显然是一个双目标的决策问题。又例如,在人才的选拔使用、招聘录用、评优评奖等活动中,由于每个人都有不同的特长、优势和劣势,组织部门总是要综合考虑各方面的情况,做出决策使得各方面的条件优势最大,劣势最小。

多目标决策问题的最主要的特点是各目标间的矛盾性和不可公度性。所谓目标间的矛盾性是指如果试图采用某一种方案去改进一个目标的值,则可能会使另一个目标的值变劣。而目标间的不可公度性是指各目标间一般没有统一的度量标准,因而不能直接进行比较。由于目标间的矛盾性和不可公度性,则不可能将多目标的问题直接归并为单目标的问题来解决。如何克服目标间的矛盾,在不可公度的目标间建立起一种可以度量的指标(即效用函数)来求解多目标决策问题,这就是本章多目标决策分析方法要研究解决的问题。

17.1 多目标决策分析的基本概念

17.1.1 多目标决策问题

1. 多目标决策问题的解决过程

多目标决策问题的求解过程主要可分为四个步骤:

第一步:问题的构成,即对所需要解决的实际问题进行分析,明确问题中的主要因素、界限和所处的环境等,从而确定问题的目标集。

第二步:建立模型,即根据第一步的结果,建立起问题的一个适宜模型。

第三步:对模型进行分析和评价,即对各种可行的方案进行比较,从而可以对每一个目标标定一个(或几个)属性(称为目标函数),这些属性的值可作为采用某方案时各个目标的一种度量。

第四步: 确定实施方案, 即依据每一个目标的属性值和预先规定的决策规则比较各可行的方案, 按优劣次序将所有的方案排序, 从而确定出最好的实施方案.

2. 多目标决策问题的基本要素

任何一个多目标决策问题都包含有五个基本要素: 决策单元、目标集、属性集、决策情况和决策规则.

决策单元是指制订决策的人, 可以是一个人, 也可以是一群人.

目标是关于决策人被研究问题的“要求”或“愿望”, 决策人可以有若干个不同的目标, 即构成一个目标集. 通常情况下, 目标集可以表示为一个递阶结构.

属性是实现目标程度的一个度量, 即每一个目标都可设定一个或若干个属性, 构成一个属性集. 目标的属性是可度量的, 它反映了特定目标达到目的的程度.

决策情况是指决策问题的结构和决策环境, 即说明决策问题的决策变量、属性, 以及度量决策变量与属性的标度、决策变量与属性之间的因果关系等.

决策规则是指用于排列方案优劣次序的规则, 而方案的优劣是依据所有目标的属性值来衡量的.

3. 多目标决策问题的数学模型

设 X 为方案集, 它是决策变量 $x = (x_1, x_2, \dots, x_N)$ 的集合, $f_1(x), f_2(x), \dots, f_n(x)$ 表示目标函数. 对于每一个给定的方案 $x \in X$, 由目标函数可以确定每一个属性 f_1, f_2, \dots, f_n 的值. 实际中, 方案集 X 可以是有限的, 也可以是无限的. 在这里我们假设决策变量 x 的所有约束都能用不等式表示出来, 即

$$g_i(x) \leq 0, i = 1, 2, \dots, m, \quad (17.1)$$

其中 $g_i(x) (i = 1, 2, \dots, m)$ 均为决策变量 x 的实值函数. 则方案集 X (又称决策空间中的可行域) 可以表示为

$$X = \{x \in \mathbb{R}^N \mid g_i(x) \leq 0, i = 1, 2, \dots, m\}.$$

于是, 一般的多目标决策问题的数学模型可以表示为

$$\begin{cases} DR[f_1(x), f_2(x), \dots, f_n(x)], \\ X = \{x \in \mathbb{R}^N \mid g_i(x) \leq 0, i = 1, 2, \dots, m\}, \end{cases} \quad (17.2)$$

其中 DR (decision rule) 表示决策规则, 即上式的意义是运用决策规则 DR 依据属性 f_1, f_2, \dots, f_n 的值在 X 中选择一个最好的方案.

例如, 设有一个确定的无限个方案的多目标决策问题, 其属性集为 $\{f_1, f_2, \dots, f_n\}$, 目标集为 $\{f_1(x), f_2(x), \dots, f_n(x)\}$, 方案集为 $X = \{x \in \mathbb{R}^N \mid g_i(x) \leq 0, i = 1, 2, \dots, m\}$, 决策规则为如果有某方案能使所有属性 f_1, f_2, \dots, f_n 都能达到最优 (最大或最小), 则可选择这个方案, 即决策问题的解. 否则, 可选择一个非劣

解的方案,使能最好地满足决策人的要求. 求解多目标决策问题的非劣解,即求向量最优化问题

$$\operatorname{opt}_{x \in X} \{f_1(x), f_2(x), \dots, f_n(x)\} \quad (17.3)$$

的解. 所得非劣解(非劣方案)集为 X^* , X^* 即可能是单元素集,也可能是多元素集. 如果是多元素集,则说明非劣解是不惟一的,决策人可根据自己的偏好选择他认为最好的方案.

选择方案的方法主要是依据决策人的“偏好结构”,一般偏好结构可用效用函数来表示,即用函数 $u(f_1(x), f_2(x), \dots, f_n(x))$ 来表示非劣方案的效用,故问题转化为求解

$$\max_{x \in X^*} u(f_1(x), f_2(x), \dots, f_n(x)) \quad (17.4)$$

的问题.

17.1.2 多属性效用函数

多属性效用函数理论是单属性效用理论的推广,效用理论可分为确定型的和不确定型的两种情形. 对确定型的效用函数通常定义为价值函数,对于更多的不确定型的情况,一般都是根据问题的实际意义来定义相应的效用函数.

1. 常用确定型的价值函数

对于确定型的效用函数通常定义为价值函数,最常用的是加性价值函数.

设 Y_1, Y_2, \dots, Y_n 分别表示 n 个属性所有可能值的集合, Y 为属性集,即 $Y = Y_1 \times Y_2 \times \dots \times Y_n$. 对于某一个确定的方案 $x = (x_1, x_2, \dots, x_n) \in X$, 对应的属性向量 $y = (y_1, y_2, \dots, y_n) \in Y$, 其中, $y_i \in Y_i$ 表示对应于方案 x 的第 i 个属性 $f_i(x)$ ($i = 1, 2, \dots, n$) 的取值. 为了简便,也称 $y = (y_1, y_2, \dots, y_n)$ 为方案,并用 v_i 表示第 i ($i = 1, 2, \dots, n$) 个属性的价值函数.

如果一个偏好结构的价值函数 $v(y)$ 能用加性表示时,则

$$v(y) = k_1 v_1(y_1) + k_2 v_2(y_2) + \dots + k_n v_n(y_n), \quad (17.5)$$

其中 k_i ($i = 1, 2, \dots, n$) 为标度常数,且 $\sum_{i=1}^n k_i = 1$.

实际上,我们可以证明:如果每一个 y_i 独立于其他的属性,则价值函数可以用加性表示,即(17.5)式成立.

进一步还可以引入价值函数的拟加性形式和乘法形式等.^[4,5]

2. 不确定型的效用函数

对于确定型的问题,一个方案只产生一个后果,而对于不确定的情况则不然. 在上一章随机性决策分析方法研究单目标决策中,我们知道,很多情况一个方案的后果不仅取决于这个方案,还依赖于自然状态,而自然状态也是不确定的,往往不受决策人控制. 现在将随机性决策分析的理论推广到多目标决策问

题,即不确定情况下的多属性效用理论.

由于后果是多属性的,类似于确定型效用函数的情况,对于决策人采取任一个行动方案的后果记为 $y = (y_1, y_2, \dots, y_n) \in Y, Y = Y_1 \times Y_2 \times \dots \times Y_n$ 为后果集. 其中 $y_i \in Y_i$ (属性集) 表示第 i ($i = 1, 2, \dots, n$) 个属性的值. 对于任一个后果 $y \in Y$ 以概率 $p(y)$ 出现,为此,我们可以定义展望 P 来表示此不确定事件. 定义展望为

$$P = (y, p(y)), y \in Y.$$

所有展望的集合记作 \bar{P} (展望集).

对于任何一个后果集 Y , 每个展望 P 都有相应的概率分布 $p(y)$ ($y \in Y$), 把与 $p(y)$ 相关联的展望记作 $P(p)$. 由于后果的多属性 $y = (y_1, y_2, \dots, y_n)$ 的不确定性, 我们可以视每一个属性为一个随机变量 y_i , 并有相应的边缘概率分布 $p_i(y_i)$ ($i = 1, 2, \dots, n$).

如果把后果 y 的效用函数记作 u , 而把 y_i 条件效用函数记作 u_i ($i = 1, 2, \dots, n$). 由此, 我们可以给出下面的定义.

定义 17.1 设 X_1, X_2, \dots, X_m 是属性集 Y 中的任何正常子集, 且 $X_1 \cup X_2 \cup \dots \cup X_m = Y$. 如果对于 \bar{P} 中的任意两个展望 $P(p)$ 和 $P(q)$, 它们对应的联合概率分布分别为 p 和 q , 并且当 $p_{x_i} = q_{x_i}$ ($i = 1, 2, \dots, m$) 时有 $P(p) \sim P(q)$ (二者无差异). 则称子集 X_1, X_2, \dots, X_m 为价值独立的.

定义中的 p_{x_i} 和 q_{x_i} 均表示子集 X_i ($i = 1, 2, \dots, m$) 上的边缘概率分布, 即说明在 X_1, X_2, \dots, X_m 上的任何两个展望无差异关系不仅依赖于其边缘分布, 而且依赖于其联合分布, 并有下面的定理.

定理 17.1^[4,5] 如果定义在 Y 上的 m 个属性的效用函数为 y_i , 对于任意的 $y', y'' \in Y$ 有

$$y' \succsim y'' \quad \text{当且仅当} \quad u(y') \geq u(y''),$$

则效用函数 $u(y)$ 可以表示为加性形式

$$u(y) = \sum_{i=1}^n k_i u_i(y_i)$$

当且仅当在属性集 Y_1, Y_2, \dots, Y_n 上价值独立性成立.

类似地, 在一定的条件下, 也可以引入效用函数拟加性形式和乘法形式.

17.2 多目标决策问题的非劣解

由于多目标决策问题客观存在着两个基本特点: 目标间的矛盾性和目标间的不可公度性. 因此, 多目标决策问题一般不存在通常意义下的最优解. 即不存在既满足约束条件, 又能使所有的属性分别达到各自的最优值. 于是, 一般情况

下,只能来寻求问题的非劣解.

17.2.1 非劣解的概念

求解多目标决策非劣解的问题,即是在决策空间的可行域 X 中求解多目标优化(或向量优化)问题(17.3)或(17.4). 问题的非劣解也称为有效解,或帕雷托(Pareto)最优解. 如果一个多目标决策问题为一多属性的极大化问题,则问题可以为

$$\begin{cases} \max_{x \in X} \{f_1(x), f_2(x), \dots, f_n(x)\}, \\ X = \{x \in \mathbf{R}^N \mid g_i(x) \leq 0, i = 1, 2, \dots, m\}. \end{cases} \quad (17.6)$$

对于决策人来说,一定喜好每个属性都有极大值的方案. 这种偏好可以为

$$x_1 > x_2 \Leftrightarrow f(x_1) \geq f(x_2) \quad (17.7)$$

这里 $f(x_1) \geq f(x_2)$ 意味着 $f_i(x_1) \geq f_i(x_2)$ ($i = 1, 2, \dots, n$), 且至少存在一个 i_0 ($1 \leq i_0 \leq n$) 使 $f_{i_0}(x_1) > f_{i_0}(x_2)$, (17.7) 式称为单调性条件. 于是,我们可以给出非劣解的定义.

定义 17.2 设 x^* 是问题(17.6)的一个可行解,如果不存在其他的可行解 $x \in X$, 使 $f(x) \geq f(x^*)$ 成立,即 $f_i(x) \geq f_i(x^*)$ ($i = 1, 2, \dots, n$), 且至少存在一个 i_0 ($1 \leq i_0 \leq n$) 使 $f_{i_0}(x) > f_{i_0}(x^*)$ 成立. 则称 x^* 为问题(17.6)的一个非劣解.

17.2.2 非劣解的求解方法

如何求多目标决策问题的非劣解呢? 在这里我们主要介绍最常用的加权法和约束法.

1. 加权法

加权法是求非劣解的一种最传统的方法,即对于向量优化问题(17.6)中的目标函数进行加权,变换为如下的数量优化问题

$$\begin{cases} \max_{x \in X} f(x, w) = \sum_{k=1}^n w_k f_k(x), \\ X = \{x \in \mathbf{R}^N \mid g_i(x) \leq 0, i = 1, 2, \dots, m\}, \end{cases} \quad (17.8)$$

亦即为一个非线性规划问题. 根据非线性规划问题的库恩-塔克(Kuhn-Tucker)条件可得可行解 x^* 为非劣解的必要条件为

$$\begin{cases} x^* \in X, \\ \mu_j g_j(x^*) = 0, j = 1, 2, \dots, m, \\ \sum_{k=1}^n w_k \nabla f_k(x^*) - \sum_{j=1}^m \mu_j \nabla g_j(x^*) = 0. \end{cases}$$

实际上,可以证明:当权值 $w_k \geq 0$ ($k = 1, 2, \dots, n$) 时,则问题(17.8)的最优

解是问题(17.6)的非劣解,但当至少有一个 $w_k = 0$ 时,则问题(17.8)的最优解不惟一,这些最优解可能会是问题(17.6)的劣解。

现在的问题是如何来选取合适的权值 $w_k (k=1,2,\cdots,n)$?

关于权值 (w_1, w_2, \cdots, w_n) 的取法可能容易同决策人对各属性的偏好联系在一起,如果用 (w_1, w_2, \cdots, w_n) 来表示决策人的偏好程度,则所得的解是最佳调和解(即决策人最有可能选用的解),但未必是问题的非劣解。因此,我们在研究非劣解时,不考虑决策人的偏好,即权值 (w_1, w_2, \cdots, w_n) 不具有任何价值意义,仅仅作为寻求非劣解的参数来考虑。

一般情况下,可以证明:对于给定的一个权向量 $w = (w_1, w_2, \cdots, w_n)$,如果数量优化问题(17.8)有最优解 x^* ,当 x^* 为(17.8)的惟一解,或 $w > 0$ 时,则 x^* 是问题(17.6)的一个非劣解。即意味着无论目标函数 f 和可行域 X 是否为凸的,只要适当地选取权向量 w 总能得到问题(17.6)的一些非劣解。相反地,如果目标函数 f 和可行域 X 为凸的,则对于任何确定的非劣解 x^* ,总能找到权向量 w 使 x^* 是问题(17.8)的最优解。换言之,如果目标函数 f 和可行域 X 的凸性条件成立,则问题(17.6)的所有非劣解都能从求解问题(17.8)得到。但是,对于给定的非劣解 x^* ,对应的权向量 w 不一定是惟一的。

2. 约束法

约束法是寻求非劣解的另一种常用的直观方法。对于多目标决策问题(17.6),首先根据决策人的偏好任意选择一个属性。例如,第 k_0 个属性 $f_{k_0}(x)$ ($1 \leq k_0 \leq n$),要求 $\max f_{k_0}(x)$,然后考虑求约束问题(线性或非线性问题)

$$\begin{cases} \max f_{k_0}(x), \\ f_k(x) \geq \varepsilon_k, k = 1, 2, \cdots, n, k \neq k_0 \end{cases} \quad (17.9)$$

的最优解。在一定的条件下,问题(17.9)的最优解就是原问题(17.6)的非劣解。其中 $\varepsilon_k (k=1,2,\cdots,n; k \neq k_0)$ 与 x 无关的常数,其要求的条件就是保证(17.9)有解。

一般情况下,求解(17.9)的问题,可以通过相应的库恩-塔克条件实现,即

$$\begin{cases} x^* \in X, \\ \mu_j g_j(x^*) = 0, j = 1, 2, \cdots, m, \\ \lambda_{k_0} \nabla f_{k_0}(x^*) + \sum_{\substack{k=1 \\ (k \neq k_0)}}^n \lambda_k \nabla f_k(x^*) - \sum_{j=1}^m \mu_j \nabla g_j(x^*) = 0. \end{cases}$$

17.3 多目标群决策问题的解

前面所研究的多目标决策问题,我们总是假设决策人是一个人,但在实际中

许多决策问题可能会涉及多个人(一群人)的利益,该人群中每一个成员都有权参与问题决策方案的选择.如何综合这个群中各成员的利益使形成群的意见,最后制订出决策方案,这是一类非常重要的群决策问题.

17.3.1 多目标群决策的数学模型

多目标群决策问题的一般数学模型如下:设有一决策群包含有 n 个决策人,其中第 i 个决策人有 m_i 个目标分别为 $f_1^{(i)}(x), f_2^{(i)}(x), \dots, f_{m_i}^{(i)}(x), x \in X$ (方案集).在这个决策问题中,第 i 个决策人的最优决策可归结为求优化问题:

$$\max_{x \in X} \{f_1^{(i)}(x), f_2^{(i)}(x), \dots, f_{m_i}^{(i)}(x)\}, \quad (17.10)$$

其中 $i=1, 2, \dots, n$.

进一步,如果假设每一个决策人都有相应的多属性效用函数

$$U_i(f_1^{(i)}(x), f_2^{(i)}(x), \dots, f_{m_i}^{(i)}(x)) \quad (i=1, 2, \dots, n),$$

则对第 i 个决策人的目标就是选择合适的方案使其效用函数达到最大,即(17.10)式等价于

$$\max_{x \in X} U_i(f_1^{(i)}(x), f_2^{(i)}(x), \dots, f_{m_i}^{(i)}(x)) \quad (i=1, 2, \dots, n).$$

对于决策群中的 n 个决策人来说都会企图使自己的效用函数达到最大,因此,群决策问题也可以表示为

$$\max_{x \in X} \{U_1(f_1^{(1)}(x), f_2^{(1)}(x), \dots, f_{m_1}^{(1)}(x)), \dots, U_n(f_1^{(n)}(x), f_2^{(n)}(x), \dots, f_{m_n}^{(n)}(x))\}. \quad (17.11)$$

如果记 m_i 维向量

$$f^{(i)}(x) = [f_1^{(i)}(x), f_2^{(i)}(x), \dots, f_{m_i}^{(i)}(x)]^T,$$

则(17.11)式就可以表示为

$$\max_{x \in X} \{U_1(f^{(1)}(x)), U_2(f^{(2)}(x)), \dots, U_n(f^{(n)}(x))\}.$$

特别地,如果决策群中的每个人的目标都是一致的,即共同的目标函数为

$$f(x) = [f_1(x), f_2(x), \dots, f_m(x)]^T,$$

则问题为

$$\max_{x \in X} \{U_1(f(x)), U_2(f(x)), \dots, U_n(f(x))\}. \quad (17.12)$$

17.3.2 多目标群决策问题的求解方法

多目标群决策问题是非常复杂的,如果不是试图寻求最佳的调和解,只求非劣解和帕雷托最优解,则问题可以作为一个两级优化问题来解决.

第一级优化是选择合适的方案使对每一个决策人都是非劣的,即不存在另一个方案使任何一个决策人的某一个目标函数值得到改进,则又不损害其他的目标函数值.

第二级优化是选择合适的方案对于决策群为帕雷托最优,即不存在另一方案对群中的某一个成员产生更大的效用,而不降低其他成员的效用。

这两级优化问题必须同时求解,实际中,具体的实现是非常困难的,主要表现在进行群的价值判断时要进行人与人之间的效用比较,这种比较通常难以实现,并随着目标数目的增多,困难也随之增加。

另一种将多目标群决策问题(17.12)的化为二级优化问题的方法可以回避上述方法中要进行人与人之间的效用比较的困难。而是首先将 n 个决策人的目标作为 n 个子系统,分别寻求决策方案使他们各自的目标达到最优,即求 n 个子系统的优化问题:

$$\max_{x \in X} U_i(f_1^{(i)}(x), f_2^{(i)}(x), \dots, f_{m_i}^{(i)}(x)) \quad (i = 1, 2, \dots, n).$$

这样所得到的 n 个决策方案 $x_i^* \in X$ ($i = 1, 2, \dots, n$) 一般不会完全重合,因此,不能形成统一的群最优决策。然后,设想在 n 个决策人之上还有一个总协调人,按照某种优化原则来协调 n 个决策人之间的矛盾。这样就形成了一个两级递阶系统的优化问题,即子系统优化级和协调优化级。目前,这种方法对于单目标的群决策问题和某些特殊的多目标群决策问题已得到了应用,而对于一般的多目标群决策问题还处于理论研究阶段。

17.3.3 同目标的群决策问题的解法

下面我们就 n 个决策人有相同的 m 个目标,其目标函数(属性)为 $y^{(j)} = f_j(x)$ ($j = 1, 2, \dots, m$),于是问题可以表示为

$$\max_{y \in Y} U_i(y^{(1)}, y^{(2)}, \dots, y^{(m)}), \quad i = 1, 2, \dots, n, \quad (17.13)$$

其中 $y = [y^{(1)}, y^{(2)}, \dots, y^{(m)}]^T$ 表示问题的属性向量, Y 为方案集 X 在目标函数空间中的映射的像集(属性集)。

下面我们按照“多数规则”来求解问题(17.13)^[4,5]。

所谓的“多数规则”:如果 $y^* \in Y$ 是问题的一个解,则不存在另外的 $y \in Y$ 使决策群中有一半以上的人认为 $y > y^*$, 所有这样的解的全体构成的集合记作 C , 称为多数核。即

$$C = \{y^* \in Y \mid \text{不存在 } y \in Y \text{ 使 } U_i(y) > U_i(y^*), \text{ 对于 } n/2 \text{ 以上的人}\}.$$

如果把 y 限制在对应的非劣解的集内,可以用 17.2 节中的约束法求非劣解,则

$$\begin{cases} \max_{x \in X} f_m(x), \\ f_i(x) \geq y^{(i)} \quad (i = 1, 2, \dots, m-1). \end{cases}$$

由此可以求得属性 $y^{(m)}$ 为其他属性 $y^{(1)}, y^{(2)}, \dots, y^{(m-1)}$ 的函数,即

$$y^{(m)} = w(y^{(1)}, y^{(2)}, \dots, y^{(m-1)}),$$

将其代入(17.13)式中有

$$\begin{cases} \max_{x \in X} U_i(y^{(1)}, y^{(2)}, \dots, y^{(m-1)}, w(y^{(1)}, y^{(2)}, \dots, y^{(m-1)})) \quad (i = 1, 2, \dots, n), \\ f_j(x) \geq y^{(j)} \quad (j = 1, 2, \dots, m-1). \end{cases} \quad (17.14)$$

如果仍记 $y = [y^{(1)}, y^{(2)}, \dots, y^{(m-1)}]^T$, 相应的属性集为 $Y^{(1)}$, 即此处的 $y \in Y^{(1)}$, 则将问题(17.14)与(17.13)相比较, 变量的个数减少了一个, 从而使问题简化了, 这样可以继续做下去. 如果我们已求得一个或多个 $y^* \in Y^{(1)}$ 和 $w(y^*)$ 使 $y^* \in C$, 即得到了问题的多数核.

实际上, 可以证明:^[4,5] 当决策人数为奇数时, 多数核只有一点; 当决策人数为偶数时, 多数核将是一个点集. 并且, 当 $m > 2$ 时, 在一定的条件下, 多目标群决策问题的多数核是存在的.

17.4 股份制公司的综合投资问题

17.4.1 问题的提出

对于股份制公司(或企业)而言, 每年都要召开股东大会, 研究确定下一年度的投资方案, 使其利用有限的资金获取最大的投资收益和最小的风险. 对于每一个股东, 将会根据自己的利益在可能的投资方案中总是选择对自己有利的方案, 董事会将综合各股东的意见和所持股份做出最后的投资决策方案. 这是一类非常有代表性的问题, 一般的问题可以描述如下:

设某股份公司(或企业)有 n 个股东, 每个股东所持股份的比例分别为 s_k ($k = 1, 2, \dots, n$). 公司计划投入 M 万元资金用于下一年度 N 个预选项目的投资, 对任何一个投资项目都是风险与收益并存, 而且风险随着收益的增加而增大. 实际中, 每个项目的收益和风险都与一些不确定的因素有关(即可视为随机变量), 其期望值分别为 p_i, r_i ($i = 1, 2, \dots, N$). 董事会规定, 如果确定投资某一个项目, 则该项目至少投资 m 万元. 现在的问题是: 董事会如何充分考虑公司的利益和综合各股东的意见做出最后的投资决策方案.

17.4.2 问题的分析与假设

根据这类问题的实际情况, 首先给出如下的假设:

- (1) 每一个投资项目的收益都是相互独立的随机变量, 记为 R_i ($i = 1, 2, \dots, N$), 即任意两个投资项目的收益互不影响, 只受某些不确定因素的影响;
- (2) 风险与收益并存, 而且高收益伴随着高风险;

(3) 每个股东都是独立的、理性的决策者,而且各股东在公司的地位受其股份的影响。

根据如上的假设,每个投资项目收益 R_i 的期望为 $E(R_i) = p_i$ ($i = 1, 2, \dots, N$), 称其为收益率;用方差 $D(R_i) = r_i$ ($i = 1, 2, \dots, N$) 表示相应的投资风险率。最佳的投资方案应该是在充分考虑公司的利益和各股东意见的基础上,使得总的投资收益最大,而相应的风险最小。

17.4.3 多目标决策模型

设第 i 个项目的投资金额占总投资金额的比率为 x_i ($i = 1, 2, \dots, N$), 则 $x_i = 0$ 或者 $\frac{m}{M} \leq x_i \leq 1$ ($i = 1, 2, \dots, N$), 其综合投资方案的总收益为 $R = \sum_{i=1}^N x_i R_i$, 总收益率为 $E(R) = \sum_{i=1}^N x_i E(R_i) = \sum_{i=1}^N x_i p_i$. 由假设(1)知 $\text{Cov}(R_i, R_j) = 0$ ($i \neq j$), 则总风险率为 $D(R) = \sum_{i=1}^N x_i D(R_i) = \sum_{i=1}^N x_i r_i$. 于是, 建立多目标决策数学模型:

$$\begin{cases} \max \sum_{i=1}^N x_i p_i, \\ \min \sum_{i=1}^N x_i r_i, \\ x_i = 0 \text{ 或 } m/M \leq x_i \leq 1 \quad (i = 1, 2, \dots, N), \\ \sum_{i=1}^N x_i = 1. \end{cases} \quad (17.15)$$

实际中, 公司对投资方案会有一个期望收益目标值 p_0 和所能承受的风险值 r_0 , 但由于受不确定因素影响, 实际方案的收益和风险与其目标值可能会有一定的偏差. 因此, 我们可以将模型转化为下面的目标规划模型:

$$\begin{cases} \min(d_1^- + \lambda d_2^+), \\ \sum_{i=1}^N x_i p_i + d_1^- - d_1^+ = p_0, \\ \sum_{i=1}^N x_i r_i + d_2^- - d_2^+ = r_0, \\ x_i = 0 \text{ 或 } m/M \leq x_i \leq 1 \quad (i = 1, 2, \dots, N), \\ \sum_{i=1}^N x_i = 1, \\ d_j^-, d_j^+ \geq 0 \text{ 且 } d_j^- \cdot d_j^+ = 0 \quad (j = 1, 2), \end{cases} \quad (17.16)$$

其中 λ 表示收益与风险之间的转化率, 即二者重要程度的量化, p_0, r_0 和 λ 都是

待定常数.

17.4.4 群决策模型

在模型(17.16)中,常数 p_0, r_0 和 λ 的确定成为求解的关键,为此需要利用群决策的方法,由董事会综合每个股东的意见和相应的股份加权确定.

实际中,每个股东对待期望收益率、风险率和二者间的转化率的看法是不同的,分别记为 p_{0k}, r_{0k} 和 λ_k ($k=1, 2, \dots, n$). 由于所持股份的差异,他们所处的地位也是不同的,因此,董事会应根据各股东的影响力适当地确定决策权值 w_k ($k=1, 2, \dots, n$),综合各股东意见的加权和得到这三个指标,即

$$p_0 = \sum_{k=1}^n w_k p_{0k}, \quad r_0 = \sum_{k=1}^n w_k r_{0k}, \quad \lambda = \sum_{k=1}^n w_k \lambda_k. \quad (17.17)$$

下面利用群决策的理论和方法来确定决策权值 w_k ($k=1, 2, \dots, n$).

首先,由每个股东根据其他股东的股份和业绩等因素给出相应的评价,即各股东之间可以相互给定一个评价权值 a_{kl} ,且 $0 \leq a_{kl} \leq 1, a_{kk} = 0, \sum_{l=1}^n a_{kl} = 1$ ($k, l=1, 2, \dots, n; k \neq l$). 一个股东的评价权值的大小反映出其他股东对该股东在决策中所起作用的信任度.

另一方面,要充分考虑股东所持股份多少的作用,于是,对股东持有股份的数量进行标准化,令 $\bar{s}_k = s_k / \sum_{l=1}^n s_l$ ($k=1, 2, \dots, n$). 一个合适的决策权值应该与其他股东对其评价权值相差不大,同时要考虑到各股东的地位,故求二者的加权偏差平方和的最小化问题

$$\begin{aligned} \min W &= \sum_{k=1}^n \sum_{l=1}^n \bar{s}_k (w_k - a_{lk})^2, \\ \text{s. t. } &\begin{cases} 0 \leq a_{lk} \leq 1, a_{kk} = 0, \sum_{k=1}^n a_{lk} = 1 \quad (k, l=1, 2, \dots, n; k \neq l), \\ w_k \geq 0 \quad (k=1, 2, \dots, n). \end{cases} \end{aligned} \quad (17.18)$$

不妨设问题(17.18)的解为 $w_k^{(0)}$ ($k=1, 2, \dots, n$),可以证明

$$w_k^{(0)} = \sum_{l=1}^n \bar{s}_l a_{lk} \quad (k=1, 2, \dots, n).$$

事实上,令

$$\frac{\partial W}{\partial w_k} = 2(w_k - \sum_{l=1}^n \bar{s}_l a_{lk}) = 0 \quad (k=1, 2, \dots, n),$$

可得 $w_k^{(0)} = \sum_{l=1}^n \bar{s}_l a_{lk}$ ($k=1, 2, \dots, n$),且对应目标函数的海赛(Hesse)矩阵是正定

的,所以 $w_k^{(0)} = \sum_{i=1}^n \bar{s}_i a_{ik}$ ($k = 1, 2, \dots, n$) 是问题(17.18)的解,即表示各股东在充分考虑持有股份的前提下其他各股东给予的客观评价权值. 一个股东的决策权值应由所持股份的比率和其他股东的评价权值两部分确定. 因此,我们定义决策权值为

$$w_k = (w_k^{(0)} \cdot \bar{s}_k) / \sum_{i=1}^n (w_i^{(0)} \cdot \bar{s}_i) \quad (k = 1, 2, \dots, n), \quad (17.19)$$

这里 $\sum_{k=1}^n w_k = 1$. 将(17.19)式代入(17.17)式得 p_0, r_0 和 λ , 再把 p_0, r_0 和 λ 代入目标规划模型(17.16)求解得组合投资比例为 $X = (x_1, x_2, \dots, x_N)$, 即为各投资项目的资金占总投资金额的比率,其投资金额为 MX .

17.4.5 模型的应用

设某上市公司董事会现有十大股东,持股比例如表 17-1.

表 17-1 十大股东的持股比例

| 股 东 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|------------|-------|------|------|------|------|-----|------|------|------|-----|
| 持股比例 s_k | 45.64 | 1.52 | 1.45 | 1.06 | 1.07 | 0.7 | 0.63 | 0.59 | 0.54 | 0.4 |

该公司计划用 8 000 万元资金用于下一年度的项目投资,初步方案有五个候选项目,其相应的期望收益率和投资风险率如表 17-2 所示.

表 17-2 投资项目的收益与风险

| 项 目 | 1 | 2 | 3 | 4 | 5 |
|-----------|------|------|------|------|------|
| 收益率 p_i | 0.12 | 0.15 | 0.18 | 0.20 | 0.25 |
| 风险率 r_i | 0.08 | 0.10 | 0.15 | 0.22 | 0.28 |

董事会规定,如果确定某个投资项目,则其项目投资额至少 100 万元.

不妨设各股东之间的相互评价权值矩阵为

$$A = \begin{bmatrix} 0 & 0.23 & 0.16 & 0.08 & 0.10 & 0.10 & 0.14 & 0.04 & 0.04 & 0.11 \\ 0.16 & 0 & 0.13 & 0.12 & 0.08 & 0.11 & 0.12 & 0.03 & 0.02 & 0.03 \\ 0.19 & 0.14 & 0 & 0.17 & 0.13 & 0.14 & 0.04 & 0.08 & 0.06 & 0.06 \\ 0.18 & 0.21 & 0.16 & 0 & 0.11 & 0.04 & 0.02 & 0.10 & 0.11 & 0.07 \\ 0.17 & 0.16 & 0.19 & 0.08 & 0 & 0.15 & 0.02 & 0.06 & 0.05 & 0.12 \\ 0.17 & 0.09 & 0.08 & 0.19 & 0.03 & 0 & 0.11 & 0.15 & 0.10 & 0.07 \\ 0.16 & 0.18 & 0.10 & 0.10 & 0.13 & 0.14 & 0 & 0.06 & 0.06 & 0.09 \\ 0.16 & 0.16 & 0.18 & 0.10 & 0.10 & 0.13 & 0.10 & 0 & 0.03 & 0.02 \\ 0.17 & 0.18 & 0.15 & 0.16 & 0.10 & 0.05 & 0.04 & 0.06 & 0 & 0.09 \\ 0.10 & 0.13 & 0.13 & 0.13 & 0.04 & 0.10 & 0.15 & 0.15 & 0.06 & 0 \end{bmatrix}.$$

由(17.18)式和(17.19)式计算得各股东的决策权值如表 17-3 所示,各股东的期望收益率、承受风险率和二者的转化率取值如表 17-4.

表 17-3 十大股东的评价权值

| 股 东 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 决策权值 w_k | 0.536 | 0.154 | 0.105 | 0.042 | 0.049 | 0.033 | 0.038 | 0.012 | 0.011 | 0.019 |

表 17-4 十大股东的收益率、风险率和转化率

| 股 东 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|-----------------|------|------|------|------|------|------|------|------|------|------|
| 收益率 p_k | 0.12 | 0.22 | 0.18 | 0.24 | 0.18 | 0.17 | 0.23 | 0.19 | 0.15 | 0.21 |
| 风险率 r_k | 0.25 | 0.08 | 0.22 | 0.16 | 0.25 | 0.18 | 0.22 | 0.17 | 0.14 | 0.12 |
| 转化率 λ_k | 3 | 5 | 3 | 4 | 3 | 5 | 4 | 5 | 5 | 4 |

由(17.17)式计算得到 $p_0 = 0.159\ 0$, $r_0 = 0.207\ 3$, $\lambda = 3.520$. 将其代入(17.16)式中求解可得该上市公司的投资比例为 $X = (0.140\ 7, 0.350\ 7, 0.294\ 2, 0.114\ 7, 0.099\ 8)$, 五个项目的投资金额为 $MX = (1\ 125, 2\ 806, 2\ 354, 917, 798)$ (万元), 即为该公司下一年度最优的组合投资方案.

由于决策方案是依据公司各股东所持股份、综合考虑了各股东的意见和地位, 利用群决策方法求得的最优组合投资方案, 求解结果符合于所有股东的利益, 使得公司的收益最大, 而风险都在可承受范围内. 该模型最突出的特点是避免了决策权被一两个大股东所控制的不公平性, 充分体现了各股东相对平等的权力和地位, 这是完全符合实际的. 该模型具有一定的应用推广价值.

17.5 参考案例与参考文献

1. 参考案例

- (1) 研究生院综合评估问题——文献[5]:236—237
- (2) 三峡工程的投资问题——文献[5]:412—417
- (3) 投入产出问题——文献[7]:179—190
- (4) 电力分配问题——文献[8]:239—242
- (5) 投资的收益和风险问题——文献[9]:327—368
- (6) 公交车调度问题——文献[10]:89—99

2. 参考文献

- [1] 杨超等. 运筹学. 北京: 科学出版社, 2004
- [2] 何坚勇. 运筹学基础. 北京: 清华大学出版社, 2000

- [3] 《运筹学》教材编写组. 运筹学. 修订版. 北京:清华大学出版社,1990
- [4] 陈 珽. 决策分析. 北京:科学出版社,1997
- [5] 岳超源. 决策理论与方法. 北京:科学出版社,2003
- [6] 叶雅阁等. 决策科学手册. 天津:天津科技翻译出版社,1989
- [7] 唐焕文等. 数学建模引论. 第二版. 北京:高等教育出版社,2001
- [8] 白其峥. 数学建模案例分析. 北京:海洋出版社,2000
- [9] 全国大学生数学建模竞赛组委会. 全国大学生数学模型竞赛优秀论文汇编. 北京:中国物价出版社,2002
- [10] 吕鹏等. 公交车调度. 工程数学学报,2002(19)建模专辑,75—80

第十八章 图论方法

图论是运筹学的一个重要分支,它广泛地应用于物理学、现代控制论、信息论、管理科学、计算机技术等诸多领域.对于自然科学、工程技术、经济管理和社会现象等诸多方面的问题,利用图论的理论和方法能够提供有力的数学模型使问题得到解决.在国内外大学生数学建模竞赛中,与图论的知识和方法有关的问题已出现多次.在这里我们仅有针对性地介绍图论的基本概念和结论以及相关的有效算法.

18.1 图

18.1.1 图的基本概念

所谓的图,直观地讲就是在平面上 n 个点,把其中的一些点对用曲线或直线连接起来,不考虑点的位置与连线曲直长短,这样形成一个关系结构就是一个图.记成 $G = (V(G), E(G))$, $V = V(G)$ 是以上述点为元素的顶点集, $E = E(G)$ 是以上述连线为元素的边集.

如果各条边都加上方向,则称为有向图,否则称为无向图.如果有有的边有方向,有的边无方向,则称为混合图.

如果任两顶点间最多有一条边,且每条边的两个端点皆不重合的图,则称为简单图.

如果图的两顶点间有边相连,则称此顶点相邻,每一对顶点都相邻的图称为完全图,否则称为非完全图,完全图记为 K_n .

若 $V(G) = X \cup Y, X \cap Y = \emptyset, |X| \cdot |Y| \neq 0$, 且 X 中无相邻的顶点对, Y 中亦然,则称图 G 为二分图;特别地,若对任意 $x \in X$, 则 x 与 Y 中每个顶点相邻,则称图 $G(V, E)$ 为完全二分图,记为 $K_{|X|, |Y|}$.

设 $v \in V(G)$, 是边 $e \in E(G)$ 的端点,则称 v 与 e 相关联,与顶点 v 关联的边数之和称为该顶点的次数,记为 $d(v)$. 可以证明 $\sum_{v \in V(G)} d(v) = 2|E(G)|$, 且由此可知:奇次顶点的总数是偶数,即所有顶点的次数之和是边数的两倍.次数为

奇数的顶点称为奇点,否则称为偶点.

设 $W = v_0 e_1 v_1 e_2 \cdots e_k v_k$, 其中 $e_i \in E(G)$, $1 \leq i \leq k$, $v_j \in V(G)$, $0 \leq j \leq k$, e_i 与 v_{i-1} 和 v_i 关联, 称 W 是图 G 的一条道路, k 为路长, v_0 为起点, v_k 为终点; 各边相异的道路称为行迹; 各顶点相异的道路称为轨道. 若 W 是一轨道, 可记为 $P(v_0, v_k)$; 起点与终点重合的道路称为回路; 起点与终点重合的轨道称为圈, 即对轨道 $P(v_0, v_k)$ 当 $v_0 = v_k$ 时成为一圈; 图中任两顶点分别为某条道路的起点与终点, 称此图为连通图.

称两顶点 u, v 分别为起点与终点的最短轨道之长为顶点 u, v 的距离; 在连通二分图 $K_{|X|, |Y|}$ 中, X 中两顶点之间的距离为偶数, X 中的顶点与 Y 中的顶点的距离为奇数. 并且可以证明下面的定理:

定理 18.1 G 为二分图的充要条件是 G 中无长为奇数的圈.

18.1.2 最短路问题

设有给定连接若干城市的公路网, 寻求从指定城市到各城市去的最短路线, 如图 18-1.

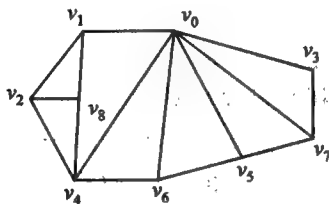


图 18-1 城市公路网示意图

建立这个问题的数学模型如下:

设任一城市为图的一个顶点 v , 连接任意两个城市的公路为图的边, 记为 e , 记 $w(e)$ 为图的边 e 之长. 对任意的顶点 $v \in V(G)$, 寻求轨道 $P(v_0, v)$, 使得

$$W(P(v_0, v)) = \min_P \{W(P)\},$$

即从 v_0 到 v 的所有轨道长中寻求最小的一个. $W(P)$ 是轨道 P 上各边长之和.

注: 若 $u, v \in V(G)$, 当 u, v 不相邻时, 则 $w(u, v) = +\infty$.

解决这一问题可以用著名的迪克斯特拉(Dijkstra)算法.

Dijkstra 算法步骤:

- (1) 令 $l(v_0) = 0, l(v) = \infty, v \neq v_0; S_0 = \{v_0\}, i = 0;$
- (2) 对每个 $v \notin S_i$, 用 $\min\{l(v), l(v_i) + w(v_i, v)\}$ 代替 $l(v)$; 设 v_{i+1} 是 $l(v)$ 使取最小值的 \bar{S}_i 中的顶点 (\bar{S}_i 是 S_i 的补集), 令 $S_{i+1} = S_i \cup \{v_{i+1}\}$ ($i = 0, 1, 2, \dots$);
- (3) 若 $i = |V(G)| - 1$, 则停止; 若 $i < |V(G)| - 1$, 令 $i = i + 1$ 转(2);

由上述的算法知, S_i 中各顶点之标志 $l(v)$ 即为 v_0 到 v 的距离, 又 $|V(G)| < \infty$, 故经有限步后 $V(G)$ 中每个顶点都标志了与 v_0 的距离, 从而可以找到 v_0 到各顶点的最短轨道. 且 Dijkstra 算法的时间复杂度(耗时量)为 $O(|V(G)|^2)$.

18.2 树

18.2.1 树的基本概念

无圈的连通图称为树, 记为 T ; 其一次顶点称为叶; 显然有边的树至少有二个叶.

若图 G 满足 $V(G) = V(T)$, $E(T) \subseteq E(G)$, 则称 T 是图 G 的生成树. 图 G 为连通的充要条件是 G 有生成树. 一个连通图的生成树不是惟一的, 用 $\tau(G)$ 表示 G 的生成树的个数, 并有下列的凯莱(Cayley)公式:

$$\tau(K_n) = n^{n-2} \quad \text{和} \quad \tau(G) = \tau(G - e) + \tau(G \cdot e),$$

其中 K_n 为 n 个顶点的完全图, $G - e$ 为从 G 中删除边 e 的图, $G \cdot e$ 为把 e 的长度收缩为零得到的图.

树有下面常用的五个充要条件:

定理 18.2 (1) G 是树当且仅当 G 中任意两顶点之间有且仅一条轨道;

(2) G 是树当且仅当 G 中无圈, 且 $|E(G)| = |V(G)| - 1$;

(3) G 是树当且仅当 G 为连通的, 且 $|E(G)| = |V(G)| - 1$;

(4) G 是树当且仅当 G 为连通的, 且对任一边 $e \in E(G)$, $G - e$ 为不连通的;

(5) G 是树当且仅当 G 中无圈, 且对任一边 $e \notin E(G)$, $G + e$ 恰有一个圈.

18.2.2 修路选线问题

假设要修建连接若干个城市的公路网, 已知 i 城与 j 城之间路的造价为 C_{ij} , 请设计一条线路使总的造价最低(如图 18-2).

这类问题的数学模型就是在连通的加权图上求权值最小的连通生成子图, 所谓图 G_1 是图 G_0 的生成子图是指 $V(G_0) = V(G_1)$, $E(G_1) \subseteq E(G_0)$. 显然, 权值最小的连通生成子图是一个生成树, 即在连通加权图上求最小的生成树. 解决这类问题可以由下面的克鲁斯卡尔(Kruskal)算法求解.

Kruskal 算法步骤:

(1) 选择边 $e_1 \in E(G)$, 使得 $w(e_1) = \min$;

(2) 若 e_1, e_2, \dots, e_i 已选好, 则从 $E(G) - \{e_1, e_2, \dots, e_i\}$ 中选取 e_{i+1} , 使得

1) $E[\{e_1, e_2, \dots, e_i, e_{i+1}\}]$ 中无圈,

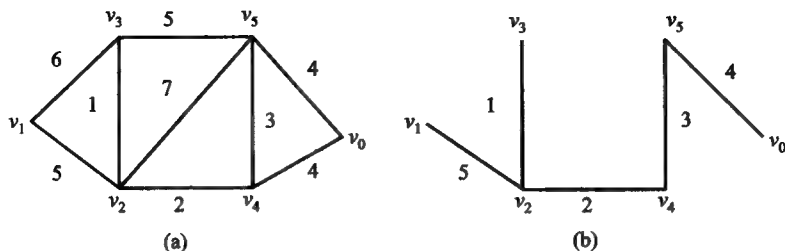


图 18-2 城市公路网示意图

2) $w(e_{i+1}) = \min;$

(3) 直到选得到 $e \mid_{V(G)} \mid_{-1}$ 为止.

其中 $E'[G](E' \subseteq E(G))$ 称为边子集 E' 的导出子图, 它是以 E' 为边集, 以 E' 中边的端点为顶点的子图.

18.3 遍历

所谓的遍历是从连通图的一个顶点出发每边恰通过一次, 或每个顶点恰通过一次. 这并不是在所有的连通图上都能实现的事情.

图中含每条边的行迹称为欧拉 (Euler) 行迹; 闭欧拉行迹称为欧拉回路; 含欧拉回路的图称为欧拉图.

直观地讲, 欧拉图就是从从一个顶点出发, 每边恰通过一次能回到出发点的图, 即不重复地行遍所有的边再回到出发点. 关于欧拉图有下面的结论.

定理 18.3 (1) G 是欧拉图的充要条件是 G 为连通图, 且每个顶点皆为偶次;

(2) G 是欧拉图的充要条件是 G 为连通图, 且 $G = \bigcup_{i=1}^k C_i$, C_i 是圈, $E(C_i) \cap E(C_j) = \emptyset$ ($i \neq j$);

(3) G 中有欧拉行迹的充要条件是 G 为连通图, 且至多有两个奇次顶点.

著名的“七桥问题”就是由欧拉证明答案是否定的, 主要是七桥图中的每个顶点都是奇次的, 由定理 18.3 的 (1) 可知, 七桥图不是欧拉图, 因此, 不可能从某一地出发, 每桥只经过一次再回到原地. 此问题可归结为一笔画问题, 即能否从某一点开始一笔画出这个图形, 最后回到原来的点, 而不重复 (如图 18-3).

图中含有所有顶点的轨道称为哈密顿 (Hamilton) 轨; 闭的哈密顿轨称为哈密顿圈; 含有哈密顿圈的图称为哈密顿图.

直观地讲, 哈密顿图就是从从一个顶点出发, 每一顶点恰通过一次能回到出发

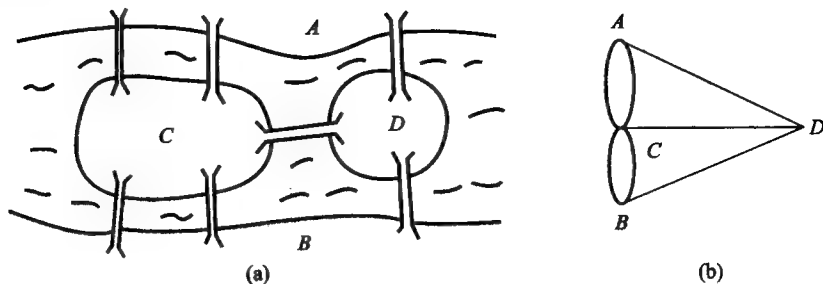


图 18-3 七桥图

点的图,即不重复地行遍所有的顶点回到出发点.

18.4 匹 配

18.4.1 匹配的概念

若 $M \subseteq E(G)$, 对任意 $e_i, e_j \in M, e_i$ 与 $e_j (i \neq j)$ 无公共端点, 则称边子集 M 为图 G 中的一个匹配; M 中一条边的两个端点称为在 M 中相配; M 中边的端点称被 M 许配; 若 G 中每个顶点皆被 M 许配, 则称 M 为完备匹配; 若 M 是 G 中的一个匹配, 但 G 中已无匹配 M' 使 $|M'| > |M|$, 则称 M 是 G 中的一个最大匹配.

若 M 是 G 中的匹配, G 中又有一个轨道, 其边交替地在 $E - M$ 与 M 中出现, 则称此轨道为 G 中 M 的交错轨; 若 M 的交错轨的起止顶点皆未被 M 许配, 则称此轨为 M 可增广轨.

M 是图 G 中最大匹配的充要条件是 G 中无 M 可增广轨.

18.4.2 分派问题

设某单位有 n 名工作人员去完成 n 项工作, 每人适合做其中一项或几项工作, 问能否每人都能分配一项合适的工作. 如果不能每人都有适合的工作, 那么最多有几个人能有适合自己的工作?

这类问题的数学模型是: 设 G 为二分图, 顶点集 $V(G)$ 可划分为 $V(G) = X \cup Y, X = \{x_1, x_2, \dots, x_n\}, Y = \{y_1, y_2, \dots, y_n\}$ 分别表示 n 个人和 n 项工作, 当且仅当 x_i 适合工作 y_j 时, $x_i y_j \in E(G)$, 现在的问题为求 G 中的最大匹配问题.

解决这类问题可用爱迪蒙斯 (Edmonds) 给出的“匈牙利算法”.

匈牙利算法步骤:

- (1) 从 G 中取一个初始匹配 M ;
- (2) 若 M 已把 X 中的顶点皆许配, 即 M 为完备匹配, 则停止; 否则, 取 X 中

未被 M 许配的一个顶点 v ,记 $S = \{v\}$, $T = \emptyset$;

(3) 若 $N(S) = T$,则无完备匹配,停止;否则取 $y \in N(S) - T$;

(4) 若 y 是被 M 许配的,则可设 $yz \in M$,用 $S \cup \{z\}$ 代替 S , $T \cup \{y\}$ 代替 T ,转(3);否则,取可增广轨 $P(v, y)$,用“对称差” $M - E(P)$ 代替 M ,转(2).

这一算法主要是把初始匹配通过可增广轨逐次增广至得到最大匹配为止.

如果要求每个人都必须分配到一份适合的工作,则这是求二分图的完全匹配问题.

18.5 图 矩 阵

18.5.1 图矩阵的概念

设有图 G ,顶点集为 $V(G)$,边集为 $E(G)$.令

$$a_{ij} = \begin{cases} 1, & v_i v_j \in E(G), \\ 0, & v_i v_j \notin E(G), \end{cases}$$

其中 $v_i, v_j \in V(G)$,则称矩阵 $A(G) = (a_{ij})_{|V(G)| \times |V(G)|}$ 为图 G 的邻接矩阵.即仅当顶点 v_i 与 v_j 之间有边 $e_k = v_i v_j$ 相连时,邻接矩阵中的元 $a_{ij} = 1$.

设

$$b_{ij} = \begin{cases} 1, & v_i \text{ 与 } e_j \text{ 相关联}, \\ 0, & v_i \text{ 与 } e_j \text{ 不关联}, \end{cases}$$

其中 $v_i \in V(G)$, $e_j \in E(G)$,则称矩阵 $B(G) = (b_{ij})_{|V(G)| \times |E(G)|}$ 为图 G 的关联矩阵(G 为无向图).即仅当顶点 v_i 是 e_j 的端点时,关联矩阵中的元 $b_{ij} = 1$.

从关联矩阵 $B(G)$ 中任删除一行得到的矩阵称为基本关联矩阵,且有下列的结论:

定理 18.4 (1) 若 G 为连通图,则关联矩阵的秩为 $r(B(G)) = |V(G)| - 1$;

(2) 若 $B_f(G)$ 是基本关联矩阵,则 G 为连通的充要条件是 $r(B_f(G)) = |V(G)| - 1$;

(3) $e_{j_1}, e_{j_2}, \dots, e_{j_r}$ ($r = |V(G)| - 1$)是图 G 的生成树的边,当且仅当这些边在 $B_f(G)$ 中对应的列构成的行列式不为0;

(4) $A^k(G)$ 中 i 行 j 列元表示 G 中从 v_i 到 v_j 的长为 k 的道路的条数.

18.5.2 非循环比赛中的实力问题

设有 n 个参赛队以非循环方式进行某种比赛,如何评价各参赛队的实力?

这类问题的数学模型为:设以 n 个参赛队 v_1, v_2, \dots, v_n 为图 G 的顶点,若参

赛队 v_i 战胜参赛队 v_j , 或战平, 则从顶点 v_i 到 v_j 连一条有向边 $v_i v_j$, 则可得一个有向图 G . 写出邻接矩阵 $A^k(G) = (a_{ij}^{(k)})_{n \times n}$, 令

$$a_i(k) = a_{i1}^{(k)} + a_{i2}^{(k)} + \cdots + a_{in}^{(k)},$$

其中 $a_{ij}^{(k)}$ 是矩阵 $A^k(G)$ 中第 i 行第 j 列元, k 为从 v_i 到 v_j 的长.

当 $a_i(k) > a_j(k)$ 时, 表明参赛队 v_i 的实力优于 v_j , 令

$$f(i) = \lim_{k \rightarrow \infty} \frac{a_i(k)}{\sum_{i=1}^n a_i(k)},$$

则 $f(i)$ 称为参赛队 v_i 的实力指数, $f(i)$ 越大, 其实力就越强.

18.6 婚配问题^[10]

现有 n 对年龄相当的青年男女, 每个人的基本条件都不相同, 每个人的择偶条件也不尽相同. 任意一对男女青年配对, 都有相应的成功率. 给出合适的配对方案, 使得:

- (1) 总的配对成功率尽可能地高;
- (2) n 对男女青年都配对成功的成功率最大.

1. 问题的分析

因为实际中任何一对男女青年都有配对成功与不成功的可能性, 所谓的“成功率”, 就是男女双方最终配对成功的概率. 对问题(1), 要使总的配对成功率尽可能地高, 也就是给出一种方案, 使得男女青年配对成功的概率之和最高. 对问题(2)要求给出一种使 n 对男女青年都能配对成功的方案, 使得这种可能性(概率)最大, 即使得每一对男女成功的概率之积最大.

2. 模型的假设

假设第 i 名男青年和第 j 名女青年的配对成功的概率为 p_{ij} , 当第 i 名男青年和第 j 名女青年能够配对时 $x_{ij} = 1$, 否则 $x_{ij} = 0$ ($i, j = 1, 2, \dots, n$).

3. 模型的建立

问题(1):

根据这一问题的特性, 将 n 对男女青年组成的集合 X 和 Y 作为二部集, 即作为二分图 $G = (V, E)$ 的顶点集, 当一对男女能够配对时, 两者之间连一条线作为二分图的边, 相互之间的配对成功的概率 p_{ij} 为对应边的权重. 于是问题就是求赋权二分图的最大匹配问题, 如图 18-4.

总配对成功的概率为

$$\max P = \sum_{i=1}^n \sum_{j=1}^n p_{ij} x_{ij}.$$

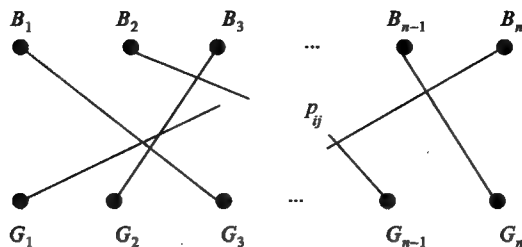


图 18-4 二分图的示意图

问题(2):

类似地,求上述的赋权二分图的最优匹配问题, n 对男女青年都配对成功的成功率可以表示为

$$\max P = \prod_{i=1}^n \prod_{j=1}^n p_{ij} x_{ij}.$$

4. 模型的求解

问题(1)可用“匈牙利算法”直接求解;而问题(2)首先对目标函数做线性化处理,即借鉴于最大似然法求极值的方法,令 $B_{ij} = \ln(p_{ij})$,然后,考虑相应的目标函数 $\max P = \sum_{i=1}^n \sum_{j=1}^n B_{ij} x_{ij}$,同样用“匈牙利算法”求解,所得的最优解也就是原问题的最优解.

例如:设有 10 对男女青年,已知配对成功率矩阵为

$$C = \begin{bmatrix} 0.832 & 0.808 & 0.804 & 0.832 & 0.884 & 0.884 & 0.788 & 0.84 & 0.84 & 0.788 \\ 0.916 & 0.644 & 0.84 & 0.884 & 1 & 0.884 & 0 & 0.936 & 0.852 & 0.82 \\ 0.744 & 0.724 & 0.772 & 0.744 & 0.756 & 0.704 & 0.704 & 0.72 & 0.636 & 0.712 \\ 1 & 0.852 & 0.804 & 1 & 0.916 & 0.872 & 0.832 & 0.936 & 0.84 & 0.84 \\ 0.936 & 0.656 & 0.772 & 0.904 & 0.884 & 0.852 & 0.644 & 0.884 & 0.832 & 0.84 \\ 0.808 & 0.676 & 0.72 & 0.688 & 0.72 & 0.72 & 0 & 0.772 & 0.788 & 0.756 \\ 0.936 & 0.644 & 0.84 & 0.936 & 0.884 & 0.84 & 0 & 0.968 & 0.84 & 0.82 \\ 0.968 & 0.704 & 0.788 & 0.916 & 0.936 & 0.884 & 0.644 & 0.884 & 0.832 & 0.884 \\ 0.916 & 0.756 & 0.72 & 0.84 & 0.84 & 0.84 & 0.676 & 0.872 & 0.852 & 0.872 \\ 0.748 & 0.688 & 0.788 & 0.768 & 0.724 & 0.656 & 0.608 & 0.724 & 0.644 & 0.704 \end{bmatrix}.$$

问题(1): 用 MATLAB 编程求解可得到配对结果为

$$e = \begin{bmatrix} 2 & 4 & 6 & 7 & 9 & 10 & 3 & 1 & 5 & 8 \\ 5 & 4 & 9 & 8 & 10 & 3 & 2 & 7 & 1 & 6 \end{bmatrix},$$

最大总成功率 $P = 8.7480$.

问题(2): 同样用 MATLAB 求解可得配对结果为

$$e = \begin{bmatrix} 2 & 4 & 6 & 7 & 5 & 9 & 8 & 1 & 10 & 3 \\ 5 & 4 & 9 & 8 & 1 & 10 & 6 & 7 & 3 & 2 \end{bmatrix},$$

其都配对的成功率为 $P=0.2474$

注: 矩阵 C 中有 0 元, 但 MATLAB 能处理对 0 取对数的情形. 因为是最大化问题, 易知 0 元的存在对最优解不会产生影响.

18.7 锁具装箱问题^{①[10]}

设一种锁具有五个槽, 每个槽有六种高度, 但要求必须有两个不同槽高, 而且相邻两个槽的高度不能为 3. 两个锁具能够互开, 当且仅当有四个对应的槽高度相同, 最后一个槽高度差 1. 销售时每 60 个锁具装一箱, 求出最大不能互开的锁具数.

首先, 我们知道满足要求的合格锁具共 5 880 种组合, 可令 $(x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5})$ 表示五个槽槽高分别为 $x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}$ 的一个锁具. 记 H_1 是所有槽高之和为奇数的锁具集合, H_2 是所有槽高之和为偶数的锁具集合. 在有奇数个槽, 每个槽有偶数高度的情况下, 集合 H_1 和 H_2 之间存在双射 $\phi: H_1 \rightarrow H_2$, 其对应关系为

$$(x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}) \mapsto (x'_{i1}, x'_{i2}, x'_{i3}, x'_{i4}, x'_{i5}),$$

其中 $x'_{ij} = 7 - x_{ij} (j=1, 2, \dots, 5)$, 所以 H_1 和 H_2 各有 2 940 个元素. 然后, 将每个合格锁具都看作一个顶点, 能够互开的两个顶点连一条边. 可计算出互开总对数为 22 778, 所以共有 22 778 条边. 因为能互开的锁具奇偶性不同, 所以 H_1 和 H_2 之间有边相连, 而 H_1, H_2 内部无边相连, 即构成一个二分图.

最大不能互开锁具数, 即为此图的最大独立顶点数. 即

$$\text{“最大独立顶点数} = \text{顶点总数} - \text{最大对集的边数”}.$$

由于此图的邻接关系比较复杂, 所以从理论上求出最大对集是很困难的, 在这里我们可以用“匈牙利算法”来求解二分图的最大独立顶点数, 即最大不能互开锁具数.

首先求出 H_1, H_2 和它们之间的邻接矩阵 W , 再用 MATLAB 求解可得到最大对集 e 及其边数 2 940. 所以最大独立顶点数 $= 5\,880 - 2\,940 = 2\,940$, 即为最大不能互开锁具数.

类似地, 也可以在两个分部顶点个数不相等的条件下寻求最大对集的问题. 这时为了更快求得最大对集, 在寻找可扩路时, 应该从顶点个数较少的分部进行寻找.

譬如: 若将“锁具装箱”的问题变为有六个槽, 每个槽有六个槽高. 此时合格

① 本案例选自 1994 年中国大学生数学建模竞赛 B 题.

锁具为 35 080 种, H_1 有 17 720 种, H_2 有 17 360 种, 互开总对数为 168 984. 然而, 邻接矩阵的维数为 $17\,720 \times 17\,360$, 规模太大, 不能用满矩阵存储, 但在 MATLAB 中可使用稀疏矩阵存储. 可求出最大对集边数为 17 360, 所以最大独立顶点数 $= 35\,080 - 17\,360 = 17\,720$.

18.8 参考案例与参考文献

1. 参考案例

- (1) 锁具装箱问题——文献[5]:161—166;[9]:57—83
- (2) 最优截断切割问题——文献[9]:276—326;[6]:134—137
- (3) 灾情巡视路线问题——文献[9]:369—409
- (4) 钢管的订购与运输问题——文献[9]:548—554
- (5) 扫雪问题——文献[7]:351—363
- (6) 投资重点的评价问题——文献[8]:75—81

2. 参考文献

- [1] 《运筹学》教材编写组. 运筹学. 修订版. 北京:清华大学出版社,1990
- [2] 杨超等. 运筹学. 北京:科学出版社,2004
- [3] 刁在筠等. 运筹学. 第二版. 北京:高等教育出版社,2001
- [4] 朱德通. 最优化模型与实验. 上海:同济大学出版社,2003
- [5] 叶其孝. 大学生数学建模竞赛辅导教材(二). 长沙:湖南教育出版社,1997
- [6] 赵静,但琦等. 数学建模与数学实验. 北京:高等教育出版社,2000
- [7] 李尚志等. 数学建模竞赛教程. 南京:江苏教育出版社,1996
- [8] 陈理荣. 数学建模导论. 北京:北京邮电大学出版社,2000
- [9] 全国大学生数学建模竞赛组委会. 全国大学生数学建模竞赛优秀论文汇编. 北京:中国物价出版社,2002
- [10] 常庭懋,韩中庚. 用“匈牙利算法”求解一类最化问题. 信息工程大学学报,2004,5(1):45—49

第十九章 模糊数学方法

模糊数学是研究和处理模糊现象的一种数学方法,它也同其他的学科一样,主要是来源于实际的需要.在社会实践中,模糊概念(或现象)无处不在.例如:在日常生活中的好与坏、大与小、厚与薄、快与慢、长与短、轻与重、高与低、贵与贱、软与硬、深与浅、美与丑、黑与白、早与晚、生与熟、动与静、穷与富、疏与密等等都包含着一定的模糊概念.随着科学技术的发展,各学科领域对与这些模糊概念有关的实际问题往往都需要给出定量的分析,因此,这就要求人们研究和处理这些模糊概念(或现象)的数学方法.

模糊数学是一个较新的现代应用数学学科,它是继经典数学、统计数学之后发展起来的一个新的数学学科.统计数学是将数学的应用范围从确定性的领域扩大到了不确定性的领域,即从必然现象到偶然现象,而模糊数学则是把数学的应用范围从确定领域扩大到了模糊领域,即从精确现象到模糊现象.我们知道,在各科学领域中,所涉及的各种量总是可以分为确定性的和不确定性的两大类,模糊数学就是研究属于不确定性,而又具有模糊性的量变化规律的一种数学方法.

实际中,我们处理现实对象的数学模型可以分为三大类:第一类是确定性的数学模型,即模型的背景具有确定性,对象之间具有必然的关系.第二类是随机性的数学模型,即模型的背景具有随机性和偶然性.第三类是模糊性模型,即模型的背景及关系都具有模糊性.我们这里所说的模糊数学建模方法就是针对实际中具有模糊性的问题,建立数学模型所需要的模糊数学的理论和知识.

19.1 模糊数学的基本概念

19.1.1 模糊集与隶属函数

1. 模糊集与隶属函数的概念

一般说来,我们对通常集合的概念并不陌生,如果将所讨论的对象限制在一定的范围内,并记所讨论的对象全体构成的集合为 U ,称之为论域,在此,总是假设问题的论域是非空的.

如果 U 是论域, 则 U 的所有子集组成的集合称为 U 的幂集, 记作 $F(U)$. 例如: $U = \{a, b, c\}$, 则 $F(U) = \{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$. 为了与模糊集相区别, 在这里称通常的集合为普通集. 对于论域 U 的每一个元素 $x \in U$ 和某一个子集 $A \subset U$, 有 $x \in A$, 或 $x \notin A$, 二者有且仅有一个成立. 于是, 对于子集 A 定义映射

$$\mu_A: U \rightarrow \{0, 1\}, \quad \text{即 } \mu_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases}$$

则称之为集合 A 的特征函数, 集合 A 可以由特征函数惟一确定.

所谓论域 U 上的模糊集 A 是指: 对任意 $x \in U$ 总以某个程度 μ_A ($\mu_A \in [0, 1]$) 属于 A , 而非 $x \in A$ 或 $x \notin A$. 也可以将普通集的特征函数的概念推广到模糊集, 即模糊集的隶属函数.

定义 19.1 设 U 是一个论域, 如果给定了一个映射

$$\mu_A: U \rightarrow [0, 1], x \mapsto \mu_A(x) \in [0, 1],$$

则就确定了一个模糊集 A , 其映射 μ_A 称为模糊集 A 的隶属函数, $\mu_A(x)$ 称为 x 对模糊集 A 的隶属度. 使 $\mu_A(x) = 0.5$ 的点 x_0 称为模糊集 A 的过渡点, 即是模糊性最大的点.

对一个确定的论域 U 可以有多个不同的模糊集, 记 U 上的模糊集的全体为 $F(U)$, 即

$$F(U) = \{A \mid \mu_A: U \rightarrow [0, 1]\},$$

则 $F(U)$ 就是论域 U 上的模糊幂集, 显然 $F(U)$ 是一个普通集合, 且 $U \subseteq F(U)$.

2. 模糊集的代表示法

对于有限的论域 $U = \{x_1, x_2, \dots, x_n\}$, A 是 U 上的任一个模糊集, 其隶属度为 $\mu_A(x_i)$ ($i = 1, 2, \dots, n$), 则模糊集的代表示形式有

(1) Zadeh 表示法

$$A = \sum_{i=1}^n \frac{\mu_A(x_i)}{x_i} = \frac{\mu_A(x_1)}{x_1} + \frac{\mu_A(x_2)}{x_2} + \dots + \frac{\mu_A(x_n)}{x_n},$$

这里 $\frac{\mu_A(x_i)}{x_i}$ 不是分数, “+”也不表示求和, 只是符号, 它表示点 x_i 对模糊集 A 的隶属度是 $\mu_A(x_i)$.

(2) 序偶表示法

$$A = \{(x_1, \mu_A(x_1)), (x_2, \mu_A(x_2)), \dots, (x_n, \mu_A(x_n))\}.$$

(3) 向量表示法

$$A = (\mu_A(x_1), \mu_A(x_2), \dots, \mu_A(x_n)).$$

对于论域 U 为无限集的情况, 则 U 上的模糊集 A 可以表示为

$A = \int_U \frac{\mu_A(x)}{x}$, 这里“ \int ”不是积分号, “ $\frac{\mu_A(x)}{x}$ ”也不是分数.

3. 模糊集的运算

模糊集与普通集有相同的运算和相应的运算规律.

定义 19.2 设模糊集 $A, B \in F(U)$, 其隶属函数为 $\mu_A(x), \mu_B(x)$.

(1) 若对任意 $x \in U$, 有 $\mu_B(x) \leq \mu_A(x)$, 则称 A 包含 B , 记 $B \subseteq A$;

(2) 若 $A \subseteq B$ 且 $B \subseteq A$, 则称 A 与 B 相等, 记为 $B = A$.

定义 19.3 设模糊集 $A, B \in F(U)$, 其隶属函数为 $\mu_A(x), \mu_B(x)$, 则称 $A \cup B$ 和 $A \cap B$ 为 A 与 B 的并集和交集; 称 A^c 为 A 的补集或余集. 它们的隶属函数分别为

$$\mu_{A \cup B}(x) = \mu_A(x) \vee \mu_B(x) = \max(\mu_A(x), \mu_B(x)),$$

$$\mu_{A \cap B}(x) = \mu_A(x) \wedge \mu_B(x) = \min(\mu_A(x), \mu_B(x)),$$

$$\mu_{A^c}(x) = 1 - \mu_A(x),$$

其中“ \vee ”和“ \wedge ”分别表示取大算子和取小算子. 并且, 并和交运算可以直接推广到任意有限的情况, 同时也满足普通集的交换律、结合律、分配律等运算^[1,2,3,4].

19.1.2 隶属函数的确定方法

我们知道, 模糊数学的基本思想是隶属程度的思想. 应用模糊数学方法建立数学模型的关键是建立符合实际的隶属函数. 然而, 如何确定一个模糊集的隶属函数至今还是尚未完全解决的问题. 这里仅介绍几种常用的确定隶属函数的方法.

1. 模糊统计方法

模糊统计方法可以算是一种客观方法, 主要是基于模糊统计试验的基础上根据隶属度的客观存在性来确定的. 所谓的模糊统计试验必须包含下面的四个要素:

(1) 论域 U ;

(2) U 中的一个固定元素 x_0 ;

(3) U 中的一个随机变动的集合 A^* (普通集);

(4) U 中的一个以 A^* 作为弹性边界的模糊集 A , 对 A^* 的变动起着制约作用. 其中 $x_0 \in A^*$, 或 $x_0 \notin A^*$, 致使 x_0 对 A 的隶属关系是不确定的.

假设我们做 n 次模糊统计试验, 则可计算出

$$x_0 \text{ 对 } A \text{ 的隶属频率} = \frac{x_0 \in A^* \text{ 的次数}}{n}.$$

事实上, 当 n 不断增大时, 隶属频率趋于稳定, 其频率的稳定值称为 x_0 对 A 的隶属度, 即

$$\mu_A(x_0) = \lim_{n \rightarrow \infty} \frac{x_0 \in A^* \text{ 的次数}}{n}.$$

2. 指派方法

指派方法是一种主观的方法,它主要依据人们的实践经验来确定某些模糊集隶属函数的一种方法.如果模糊集定义在实数域 \mathbf{R} 上,则模糊集的隶属函数称为模糊分布.所谓的指派方法就是根据问题的性质主观地选用某些形式的模糊分布,再依据实际测量数据确定其中所包含的参数.常用的模糊分布如表 19-1.

表 19-1 常用的模糊分布

| | (a) 偏小型 | (b) 中间型 | (c) 偏大型 |
|--------------|--|---|--|
| 矩形分布 | $\mu_A(x) = \begin{cases} 1, & x \leq a, \\ 0, & x > a \end{cases}$ | $\mu_A(x) = \begin{cases} 1, & a \leq x \leq b, \\ 0, & x < a \text{ 或 } x > b \end{cases}$ | $\mu_A(x) = \begin{cases} 1, & x \geq a, \\ 0, & x < a \end{cases}$ |
| 梯形分布 | $\mu_A(x) = \begin{cases} 1, & x < a, \\ \frac{b-x}{b-a}, & a \leq x \leq b, \\ 0, & x > b \end{cases}$ | $\mu_A(x) = \begin{cases} \frac{x-a}{b-a}, & a \leq x < b, \\ 1, & b \leq x < c, \\ \frac{d-x}{d-c}, & c \leq x < d, \\ 0, & x < a, x \geq d \end{cases}$ | $\mu_A(x) = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x \leq b, \\ 1, & x > b \end{cases}$ |
| 正态分布 | $\mu_A(x) = \begin{cases} 1, & x \leq a, \\ e^{-\left(\frac{x-a}{\sigma}\right)^2}, & x > a \end{cases}$ | $\mu_A(x) = e^{-\left(\frac{x-a}{\sigma}\right)^2}$ | $\mu_A(x) = \begin{cases} 0, & x \leq a, \\ 1 - e^{-\left(\frac{x-a}{\sigma}\right)^2}, & x > a \end{cases}$ |
| k 次抛物线型分布 | $\mu_A(x) = \begin{cases} 1, & x < a, \\ \left(\frac{b-x}{b-a}\right)^k, & a \leq x \leq b, \\ 0, & x > b \end{cases}$ | $\mu_A(x) = \begin{cases} \left(\frac{x-a}{b-a}\right)^k, & a \leq x < b, \\ 1, & b \leq x < c, \\ \left(\frac{d-x}{d-c}\right)^k, & c \leq x < d, \\ 0, & x < a, x \geq d \end{cases}$ | $\mu_A(x) = \begin{cases} 0, & x < a, \\ \left(\frac{x-a}{b-a}\right)^k, & a \leq x \leq b, \\ 1, & x > b \end{cases}$ |
| Γ 型分布 | $\mu_A(x) = \begin{cases} 1, & x < a \\ (k > 0), \\ e^{-k(x-a)}, & x \geq a \end{cases}$ | $\mu_A(x) = \begin{cases} e^{k(x-a)}, & x < a, \\ 1, & a \leq x < b \\ (k > 0), \\ e^{-k(x-b)}, & x \geq b \end{cases}$ | $\mu_A(x) = \begin{cases} 0, & x < a \\ (k > 0), \\ 1 - e^{-k(x-a)}, & x \geq a \end{cases}$ |
| 柯西型分布 | $\mu_A(x) = \begin{cases} 1, & x \leq a, \\ \frac{1}{1 + \alpha(x-a)^\beta}, & x > a \\ (\alpha > 0, \beta > 0) \end{cases}$ | $\mu_A(x) = \frac{1}{1 + \alpha(x-a)^\beta}$ ($\alpha > 0, \beta > 0$ 且为偶数) | $\mu_A(x) = \begin{cases} 0, & x \leq a, \\ \frac{1}{1 + \alpha(x-a)^\beta}, & x > a \\ (\alpha > 0, \beta > 0) \end{cases}$ |

实际中,根据问题对研究对象的描述来选择适当的模糊分布.偏小型模糊分布一般适合于描述像“小”、“少”、“浅”、“淡”、“冷”、“疏”、“青年”等偏向小的程度的模糊现象.偏大型模糊分布一般适合于描述像“大”、“多”、“深”、“浓”、“热”、“密”、“老年”等偏向大的程度的模糊现象.而中间型模糊分布一般适合于描述像“中”、“适中”、“不太多”、“不太少”、“不太深”、“不太浓”、“暖和”、“中年”等处于中间状态的模糊现象.但这些方法所给出的隶属函数都是近似的,应用时需要实际问题进行分析,逐步地进行修改完善,最后得到近似程度更好的隶属函数.

3. 其他方法

实际中,用来确定模糊集的隶属函数的方法是多种多样的,主要是根据问题的实际意义来确定.譬如,在经济管理、社会管理中,可以直接借助于已有的“客观尺度”作为模糊集的隶属度.如果论域 U 表示机器设备,在 U 上定义模糊集 A = “设备完好”,则可以用“设备完好率”作为 A 的隶属度.如果 U 表示产品,在 U 上定义模糊集 A = “质量稳定”,则可用产品的“正品率”作为 A 的隶属度.如果 U 表示家庭,在 U 上定义模糊集 A = “家庭贫困”,则可以用 Engel 系数 = $\frac{\text{食品消费}}{\text{总消费}}$ 作为 A 的隶属度.

另外,对于有些模糊集而言,直接给出隶属度有时是很困难的,但可以利用所谓的“二元对比排序法”来确定,即首先通过两两比较确定两个元素相应隶属度的大小排出顺序,然后用数学方法加工处理得到所需要的隶属函数.

19.2 模糊关系与模糊矩阵

19.2.1 模糊关系与模糊矩阵的概念

定义 19.4 设论域 U, V , 则称乘积空间 $U \times V$ 上的一个模糊子集 $\tilde{R} \in F(U \times V)$ 为从 U 到 V 的模糊关系. 如果 \tilde{R} 的隶属函数为

$$\mu_{\tilde{R}}: U \times V \rightarrow [0, 1], (x, y) \mapsto \mu_{\tilde{R}}(x, y),$$

则称隶属度 $\mu_{\tilde{R}}(x, y)$ 为 (x, y) 关于模糊关系 \tilde{R} 的相关程度.

由于模糊关系就是乘积空间 $U \times V$ 上的一个模糊子集,因此,模糊关系同样具有模糊集的运算及性质.

设 $U = \{x_1, x_2, \dots, x_m\}, V = \{y_1, y_2, \dots, y_n\}$, \tilde{R} 是由 U 到 V 的模糊关系,其隶属函数为 $\mu_{\tilde{R}}(x, y)$, 对任意的 $(x_i, y_j) \in U \times V$ 有 $\mu_{\tilde{R}}(x_i, y_j) = r_{ij} \in [0, 1]$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$), 记 $R = (r_{ij})_{m \times n}$, 则 R 就是所谓的模糊矩阵,于是有下面的一般性定义.

定义 19.5 设矩阵 $R = (r_{ij})_{m \times n}$, 且 $r_{ij} \in [0, 1]$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$), 则称 R 为模糊矩阵.

特别地, 如果 $r_{ij} \in \{0, 1\}$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$), 则称 R 为布尔 (Bool) 矩阵.

当 $m = 1$, 或 $n = 1$ 时, 则相应的模糊矩阵为 $R = (r_1, r_2, \dots, r_n)$, 或 $R = (r_1, r_2, \dots, r_m)^T$, 则分别称为模糊行向量和模糊列向量.

19.2.2 模糊等价与模糊相似

定义 19.6 若模糊关系 $\tilde{R} \in F(U \times V)$, 且满足

- (1) 自反性: $\mu_{\tilde{R}}(x, x) = 1$;
- (2) 对称性: $\mu_{\tilde{R}}(x, y) = \mu_{\tilde{R}}(y, x)$;
- (3) 传递性: $\tilde{R} \circ \tilde{R} \subseteq \tilde{R}$ (或 $\mu_{\tilde{R} \circ \tilde{R}}(x, y) = \bigvee_{z \in V} (\mu_{\tilde{R}}(x, z) \wedge \mu_{\tilde{R}}(z, y)) \leq \mu_{\tilde{R}}(x, y)$).

则称 \tilde{R} 是 U 上的一个模糊等价关系, 其隶属度函数 $\mu_{\tilde{R}}(x, y)$ 表示 (x, y) 的相关程度.

当论域为 $U = \{x_1, x_2, \dots, x_n\}$ 时, U 上的模糊等价关系可表示为 $n \times n$ 阶模糊等价矩阵 $R = (r_{ij})_{n \times n}$.

定义 19.7 设论域为 $U = \{x_1, x_2, \dots, x_n\}$, I 为单位矩阵, 如果模糊矩阵 $R = (r_{ij})_{n \times n}$ 满足

- (1) 自反性: $I \leq R$ (或 $r_{ii} = 1, i = 1, 2, \dots, n$);
- (2) 对称性: $R^T = R$ (或 $r_{ij} = r_{ji}; i, j = 1, 2, \dots, n$);
- (3) 传递性: $R \circ R \leq R$ (或 $\bigvee_{k=1}^n (r_{ik} \wedge r_{kj}) \leq r_{ij}; i, j = 1, 2, \dots, n$).

则称 R 为模糊等价矩阵.

实际中, 要建立一个模糊等价关系或模糊等价矩阵往往是困难的, 这主要是由于传递性难以满足. 但是, 对于满足自反性和对称性的模糊关系 \tilde{R} 与模糊矩阵 R , 则分别称为模糊相似关系与模糊相似矩阵.

19.2.3 λ 截矩阵与传递矩阵

定义 19.8 设 $R = (r_{ij})_{m \times n}$ 为模糊矩阵, 对任意的 $\lambda \in [0, 1]$.

(1) 如果令

$$r_{ij}(\lambda) = \begin{cases} 1, & r_{ij} \geq \lambda, \\ 0, & r_{ij} < \lambda \end{cases} \quad \left(\begin{matrix} i = 1, 2, \dots, m; \\ j = 1, 2, \dots, n \end{matrix} \right),$$

则称 $R_\lambda = (r_{ij}(\lambda))_{m \times n}$ 为 R 的 λ 截矩阵.

(2) 如果令

$$r_{ij}(\lambda) = \begin{cases} 1, & r_{ij} > \lambda, \\ 0, & r_{ij} \leq \lambda \end{cases} \begin{pmatrix} i = 1, 2, \dots, m; \\ j = 1, 2, \dots, n \end{pmatrix},$$

则称 $R_\lambda = (r_{ij}(\lambda))_{m \times n}$ 为 R 的 λ 强截矩阵.

显然, 对任意的 $\lambda \in [0, 1]$, λ 截矩阵是布尔矩阵.

定义 19.9 设 R 是 $n \times n$ 阶的模糊矩阵, 如果满足

$$R \circ R = R^2 \leq R \text{ (或 } \bigvee_{k=1}^n (r_{ik} \wedge r_{kj}) \leq r_{ij}; i, j = 1, 2, \dots, n),$$

则称 R 为模糊传递矩阵. 将包含 R 的最小的模糊传递矩阵称为 R 的传递包, 记为 $t(R)$.

事实上, 对于任意的模糊矩阵 $R = (r_{ij})_{n \times n}$, 则 $t(R) = \bigcup_{k=1}^n R^k = (\bigvee_{k=1}^n r_{ij}^{(k)})_{n \times n}$. 特别地, 当 R 为模糊相似矩阵时, 则存在一个最小的自然数 k ($k \leq n$), 使得 $t(R) = R^k$, 对任意自然数 $l > k$ 都有 $R^l = R^k$, 此时 $t(R)$ 一定为模糊等价矩阵.

19.3 模糊聚类分析方法

在许多工程技术和经济管理, 常常需要对某些指标按一定的标准(相似的程度、亲疏关系等)进行分类处理. 例如, 根据生物的某些性态对其进行分类、根据空气的性质对空气质量进行分类, 以及工业上对产品质量的分类、工程上对工程规模的分类、图像识别中对图形的分类、地质学中对地质土壤的分类、水资源中的水质分类等等. 这种对客观事物按一定标准进行分类的数学方法主要就是聚类分析法, 而模糊聚类分析法就是根据事物的某些模糊性质进行分类的一种数学方法. 下面给出模糊聚类分析方法的一般步骤.

19.3.1 数据标准化

(1) 获取数据: 设论域 $U = \{x_1, x_2, \dots, x_n\}$ 为所需分类研究的对象, 每个对象又由 m 个指标表示其性态, 即 $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$ ($i = 1, 2, \dots, n$), 于是, 可以得问题的原始数据矩阵为 $A = (x_{ij})_{n \times m}$.

(2) 数据的标准化处理: 在实际问题中的数据可能有不同的性质和不同的量纲, 为了使原始数据能够适合模糊聚类的要求, 需要将原始数据矩阵 A 做标准化的处理, 即通过适当的数据变换和压缩, 将其转化为模糊矩阵. 常用的方法

有以下两种:

(i) 平移—标准差变换

如果原始数据之间有不同的量纲,则可以采用这种变换后使每个变量的均值为0,标准差为1,即可以消除量纲的差异的影响.即令

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (i = 1, 2, \dots, n; j = 1, 2, \dots, m),$$

其中 $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$, $s_j = \left[\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \right]^{\frac{1}{2}}$ ($j = 1, 2, \dots, m$).

(ii) 平移—极差变换

如果经过平移—标准差变换后还有某些 $x'_{ij} \notin [0, 1]$, 则还需对其进行平移—极差变换, 即令

$$x''_{ij} = \frac{x'_{ij} - \min_{1 \leq i \leq n} \{x'_{ij}\}}{\max_{1 \leq i \leq n} \{x'_{ij}\} - \min_{1 \leq i \leq n} \{x'_{ij}\}} \quad (j = 1, 2, \dots, m).$$

显然所有的 $x''_{ij} \in [0, 1]$, 且也不存在量纲因素的影响, 从而可以得到模糊矩阵 $R = (x''_{ij})_{n \times m}$.

19.3.2 建立模糊相似矩阵

设论域 $U = \{x_1, x_2, \dots, x_n\}$, $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$ ($i = 1, 2, \dots, n$), 即数据矩阵为 $A = (x_{ij})_{n \times m}$, 如果 x_i 与 x_j 的相似程度为 $r_{ij} = R(x_i, x_j)$ ($i, j = 1, 2, \dots, n$), 则称之为相似系数, 确定相似系数 r_{ij} 有多种不同的方法.

(1) 数量积法: 对于 $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\} \in U$, 令 $M = \max_{i \neq j} \left(\sum_{k=1}^m x_{ik} \cdot x_{jk} \right)$, 则

取 $r_{ij} = \begin{cases} 1, & i = j, \\ \frac{1}{M} \sum_{k=1}^m x_{ik} \cdot x_{jk}, & i \neq j, \end{cases}$ 显然 $|r_{ij}| \in [0, 1]$. 若出现有某些 $r_{ij} < 0$, 可令

$r'_{ij} = \frac{r_{ij} + 1}{2}$, 则有 $r'_{ij} \in [0, 1]$. 也可以用平移—极差变换将其压缩到 $[0, 1]$ 上,

即可以得到模糊相似矩阵 $R = (r_{ij})_{n \times n}$.

(2) 夹角余弦法: 令

$$r_{ij} = \frac{\left| \sum_{k=1}^m x_{ik} \cdot x_{jk} \right|}{\sqrt{\sum_{k=1}^m x_{ik}^2 \cdot \sum_{k=1}^m x_{jk}^2}} \quad (i, j = 1, 2, \dots, n),$$

则 $R = (r_{ij})_{n \times n}$.

(3) 相关系数法: 令

$$r_{ij} = \frac{\left| \sum_{k=1}^m (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j) \right|}{\sqrt{\sum_{k=1}^m (x_{ik} - \bar{x}_i)^2 \cdot \sum_{k=1}^m (x_{jk} - \bar{x}_j)^2}} \quad (i, j = 1, 2, \dots, n),$$

其中 $\bar{x}_i = \frac{1}{m} \sum_{k=1}^m x_{ik}$, $\bar{x}_j = \frac{1}{m} \sum_{k=1}^m x_{jk}$, 则 $R = (r_{ij})_{n \times n}$.

注: $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$ 中的样本 x_{ik} 属于同一个样本空间 $X_i (k = 1, 2, \dots, m)$.

(4) 指数相似系数法: 令

$$r_{ij} = \frac{1}{m} \sum_{k=1}^m \exp \left\{ -\frac{3}{4} \frac{(x_{ik} - x_{jk})^2}{s_k^2} \right\},$$

其中 $s_k = \frac{1}{n} \sum_{i=1}^n (x_{ik} - \bar{x}_k)^2$, $\bar{x}_k = \frac{1}{n} \sum_{i=1}^n x_{ik} (k = 1, 2, \dots, m)$. 则 $R = (r_{ij})_{n \times n}$.

注: $x_i = \{x_{i1}, x_{i2}, \dots, x_{im}\}$ 中的样本 x_{ik} 属于不同的样本空间 X_k , 即 $x_{ik} \in X_k (k = 1, 2, \dots, m)$.

(5) 最大最小值法: 令

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} \wedge x_{jk})}{\sum_{k=1}^m (x_{ik} \vee x_{jk})} \quad (x_{ij} > 0; i, j = 1, 2, \dots, n),$$

则 $R = (r_{ij})_{n \times n}$.

(6) 算术平均值法: 令

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} \wedge x_{jk})}{\frac{1}{2} \sum_{k=1}^m (x_{ik} + x_{jk})} \quad (x_{ij} > 0; i, j = 1, 2, \dots, n),$$

则 $R = (r_{ij})_{n \times n}$.

(7) 几何平均值法: 令

$$r_{ij} = \frac{\sum_{k=1}^m (x_{ik} \wedge x_{jk})}{\sum_{k=1}^m \sqrt{x_{ik} \cdot x_{jk}}} \quad (x_{ij} > 0; i, j = 1, 2, \dots, n),$$

则 $R = (r_{ij})_{n \times n}$.

(8) 绝对值倒数法: 令

$$r_{ij} = \begin{cases} 1, & i = j, \\ M \left(\sum_{k=1}^m |x_{ik} - x_{jk}| \right)^{-1}, & i \neq j, \end{cases}$$

其中 M 为使得所有 $r_{ij} \in [0, 1]$ ($i, j = 1, 2, \dots, n$) 的确定常数, 则 $R = (r_{ij})_{n \times n}$.

(9) 绝对值指数法: 令

$$r_{ij} = \exp \left\{ - \sum_{k=1}^m |x_{ik} - x_{jk}| \right\} \quad (i, j = 1, 2, \dots, n),$$

则 $R = (r_{ij})_{n \times n}$.

(10) 海明距离法: 令

$$\begin{cases} r_{ij} = 1 - H \cdot d(x_i, x_j), \\ d(x_i, x_j) = \sum_{k=1}^m |x_{ik} - x_{jk}| \end{cases} \quad (i, j = 1, 2, \dots, n),$$

其中 H 为使得所有 $r_{ij} \in [0, 1]$ ($i, j = 1, 2, \dots, n$) 的确定常数. 则 $R = (r_{ij})_{n \times n}$.

(11) 欧氏距离法: 令

$$\begin{cases} r_{ij} = 1 - E \cdot d(x_i, x_j), \\ d(x_i, x_j) = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \end{cases} \quad (i, j = 1, 2, \dots, n),$$

其中 E 为使得所有 $r_{ij} \in [0, 1]$ ($i, j = 1, 2, \dots, n$) 的确定常数. 则 $R = (r_{ij})_{n \times n}$.

(12) 切比雪夫距离法: 令

$$\begin{cases} r_{ij} = 1 - Q \cdot d(x_i, x_j), \\ d(x_i, x_j) = \bigvee_{k=1}^m |x_{ik} - x_{jk}| \end{cases} \quad (i, j = 1, 2, \dots, n),$$

其中 Q 为使得所有 $r_{ij} \in [0, 1]$ ($i, j = 1, 2, \dots, n$) 的确定常数. 则 $R = (r_{ij})_{n \times n}$.

(13) 主观评分法: 设有 N 个专家组成专家组 $\{p_1, p_2, \dots, p_N\}$, 让每一位专家对所研究的对象 x_i 与 x_j 相似程度给出评价, 并对自己的自信度作出评估. 如果第 k 位专家 p_k 关于对象 x_i 与 x_j 的相似度评价为 $r_{ij}(k)$, 对自己的自信度评估为 $a_{ij}(k)$ ($i, j = 1, 2, \dots, n$), 则相关系数定义为

$$r_{ij} = \frac{\sum_{k=1}^N (a_{ij}(k) \cdot r_{ij}(k))}{\sum_{k=1}^N a_{ij}(k)} \quad (i, j = 1, 2, \dots, n).$$

则 $R = (r_{ij})_{n \times n}$.

综上所述,以上给出了实际中能够使用的一些方法,具体地选择要根据具体问题的性质和使用的方便来确定.

19.3.3 聚类

所谓的聚类方法就是依据模糊矩阵将所研究的对象进行分类的方法,对于不同的置信水平 $\lambda \in [0, 1]$,可得不同的分类结果,从而可以形成动态聚类图.常用的方法可以分为两类,一类是基于模糊等价矩阵的聚类方法,另一类是直接聚类方法.

(1) 传递闭包法

用上节的方法所建立的模糊矩阵 R 一般只是一个模糊相似矩阵,即 R 不一定是模糊等价矩阵.为此,首先需要由 R 来构造一个模糊等价矩阵 R^* ,根据传递闭包的性质,可以用平方法求出 R 的传递包 $t(R) = R^*$,即为一模糊等价矩阵.然后,由大到小取一组 $\lambda \in [0, 1]$ 值,确定相应的 λ 截矩阵,则可以将其分类,同时也可构成动态聚类图^[1].

(2) 布尔矩阵法

设论域为 $U = \{x_1, x_2, \dots, x_n\}$, R 是 U 上的模糊相似矩阵,对于确定的 λ 水平要求 U 中元素的分类.首先,由于模糊相似矩阵 R 作出其 λ 截矩阵 $R_\lambda = (r_{ij}(\lambda))$,即 R_λ 为布尔矩阵.然后,依据 R_λ 中的 1 元素可以将其分类.

如果 R_λ 为等价矩阵,则 R 也为等价矩阵,即可以直接将其分类.

如果 R_λ 不是等价矩阵,则首先按一定的规则将 R_λ 改造成一个等价的布尔矩阵,然后再进行分类.例如:0 元素和 1 元素互换方法等.

(3) 直接聚类法

所谓直接聚类法是一种直接由模糊相似矩阵求出聚类图的方法,具体步骤如下:

1) 取 $\lambda_1 = 1$ (最大值),对每个 x_i 作相似类: $[x_i]_R = \{x_j \mid r_{ij} = 1\}$,即将满足 $r_{ij} = 1$ 的 x_i 与 x_j 视为一类,构成相似类.相似类与等价类有所不同,不同的相似类可能有公共元素,即可能有 $[x_i]_R \cap [x_j]_R \neq \emptyset$,实际中,对于这种情况可以将 $[x_i]_R$ 与 $[x_j]_R$ 合并为一类,即可得到 $\lambda_1 = 1$ 水平上的等价分类.

2) 取 $\lambda_2 (\lambda_2 < \lambda_1)$ 为次大值,从 R 中直接找出相似程度为 λ_2 的元素对 (x_i, x_j) (即 $r_{ij} = \lambda_2$),并相应地将对应于 $\lambda_1 = 1$ 的等价分类中 x_i 与 x_j 所在的类合并

为一类,即可得到 λ_2 水平上的等价分类.

3) 依次取 $\lambda_1 > \lambda_2 > \lambda_3 > \cdots$,按第2)步的方法依次类推,直到合并到 U 成为一类为止,最后可以得到动态聚类图.

19.4 模糊模式识别方法

将事物的整体划分为若干类型而得到一组标准模式,对于一个确定的对象识别它属于哪一类的问题称为模式识别.如果整体被划分的类型与被识别的对象之中至少有一个是用模糊集表示的模式识别问题,则称为模糊模式识别.实际中有很多问题都属于这一类问题,例如:像自动分拣机对信件上邮政编码的识别;医生针对病人的主要症状诊断过程;根据学生的德、智、体等因素对学生进行分类;对某种产品等级的分类,以及指纹识别和汽车车牌号码的识别等问题.这里主要介绍两种最基本的模糊模式识别方法——最大隶属原则和择近原则.

19.4.1 模式识别中的最大隶属原则

定义 19.10 设论域 $U = \{x_1, x_2, \cdots, x_n\}$ 上的 m 个模糊子集 A_1, A_2, \cdots, A_m ,其隶属度函数为 $\mu_{A_i}(x)$ ($i = 1, 2, \cdots, m$),而模糊向量集合族 $A = (A_1, A_2, \cdots, A_m)$ 对于普通向量 $x^{(0)} = (x_1^{(0)}, x_2^{(0)}, \cdots, x_m^{(0)})$,则称 $\mu(x^{(0)}) = \bigwedge_{i=1}^m \{\mu_{A_i}(x_i^{(0)})\}$ 为 $x^{(0)}$ 对模糊向量集合族 A 的隶属度.

实际中向量 $x^{(0)}$ 对模糊向量集合族 A 的隶属度也可以定义为 $\mu(x^{(0)}) = \frac{1}{n} \sum_{i=1}^m \mu_{A_i}(x_i^{(0)})$.

1. 最大隶属原则 I

设在论域 $U = \{x_1, x_2, \cdots, x_n\}$ 上有 m 个模糊子集 A_1, A_2, \cdots, A_m (即 m 个模式)一起构成一个标准模式库,若对任一个 $x^{(0)} \in U$,存在 k_0 ($1 \leq k_0 \leq m$)使得 $\mu_{k_0}(x^{(0)}) = \bigvee_{k=1}^m \{\mu_{A_k}(x^{(0)})\}$,则可视为 $x^{(0)}$ 相对隶属于 A_{k_0} .

2. 最大隶属原则 II

设在论域 $U = \{x_1, x_2, \cdots, x_n\}$ 上确定一个标准模式 A_0 ,对于 n 个待识别的对象 $x_1, x_2, \cdots, x_n \in U$,如果有某个 x_k 满足 $\mu_{A_0}(x_k) = \bigvee_{i=1}^m \{\mu_{A_0}(x_i)\}$ ($1 \leq k \leq n$),则 x_k 优先隶属于 A_0 .

19.4.2 模式识别中的择近原则

设论域 $U = \{x_1, x_2, \cdots, x_n\}$,由 U 上的 m 个模糊子集 A_1, A_2, \cdots, A_m (即 m 个

模式) 构成一个标准模式库, 对 U 上的另一个模糊子集 A_0 , 问题是 A_0 与 $A_i (i = 1, 2, \dots, m)$ 中的哪一个最贴近? 这是另一类模式识别问题, 主要是研究两个模糊集的贴近程度.

1. 贴近度的概念

设论域 U 上的模糊子集 $A_1, A_2 \in F(U)$, 则定义

$$A_1 \circ A_2 = \bigvee_{x \in U} (\mu_{A_1}(x) \wedge \mu_{A_2}(x))$$

为 A_1 与 A_2 的内积; 类似的定义

$$A_1 \otimes A_2 = \bigwedge_{x \in U} (\mu_{A_1}(x) \vee \mu_{A_2}(x))$$

为 A_1 与 A_2 的外积.

定义 19.11 设有论域 U 上的模糊子集 $A_1, A_2 \in F(U)$, 则称

$$N(A_1, A_2) = \frac{1}{2} [A_1 \circ A_2 + (1 - A_1 \otimes A_2)]$$

为 A_1 与 A_2 的贴近度.

显然, 如果 A_1 与 A_2 的贴近度 $N(A_1, A_2)$ 越大, 则说明 A_1 与 A_2 越贴近. 而且贴近度有下列性质:

- (1) $0 \leq N(A_1, A_2) \leq 1$;
- (2) $N(U, \emptyset) = 0, N(A, A) = 1 (\forall A \in F(U))$.

实际中, 可以用贴近度来描述模糊集之间的贴近程度, 但是, 根据所研究问题的性质, 还可以给出其他形式的贴近度定义.

2. 单个特性的择近原则

设论域 U 上的 m 个模糊子集 A_1, A_2, \dots, A_m (m 个模式) 构成一个标准模式库 $\{A_1, A_2, \dots, A_m\}$, 模糊子集 A_0 为待识别的模式, 若存在 $k_0 (1 \leq k_0 \leq m)$ 使得

$$N(A_{k_0}, A_0) = \bigvee_{k=1}^m N(A_k, A_0),$$

则 A_0 与 A_{k_0} 最贴近, 或者说把 A_0 可归并到 A_{k_0} 类.

3. 多个特性的择近原则

根据实际问题的需要, 依据对象的多个特性的模式识别问题, 即要研究两个模糊向量集合族的贴近度问题, 可以有多种不同的定义, 常用的有以下几种形式:

对于论域 U 上的两个模糊向量集合族 $A = (A_1, A_2, \dots, A_m), B = (B_1, B_2, \dots, B_m)$, 则 A 与 B 的贴近度可定义为

$$(1) N(A, B) = \bigwedge_{k=1}^m N(A_k, B_k);$$

$$(2) N(A, B) = \bigvee_{k=1}^m N(A_k, B_k);$$

$$(3) N(A, B) = \sum_{k=1}^m a_k N(A_k, B_k), \text{ 其中 } a_k \in [0, 1], \text{ 且 } \sum_{k=1}^m a_k = 1;$$

$$(4) N(A, B) = \bigvee_{k=1}^m (a_k \cdot N(A_k, B_k)), \text{ 其中 } a_k \in [0, 1], \text{ 且 } \sum_{k=1}^m a_k = 1;$$

$$(5) N(A, B) = \bigvee_{k=1}^m (a_k \wedge N(A_k, B_k)), \text{ 其中 } a_k \in [0, 1], \text{ 且 } \sum_{k=1}^m a_k = 1.$$

实际中,选择哪一种形式,完全根据实际问题的需要确定,也可以用其他更合适的形式.

多个特性的择近原则:设由论域 U 上的 n 个模糊子集 A_1, A_2, \dots, A_n 构成一个标准模式库 $\{A_1, A_2, \dots, A_n\}$, 每个模式 A_k 都可用 m 个特性描述,即

$$A_k = (A_{k1}, A_{k2}, \dots, A_{km}) (k = 1, 2, \dots, n).$$

待识别的模式为 $A_0 = (A_{01}, A_{02}, \dots, A_{0m})$. 如果两个模糊向量集合族的贴近度最小值为

$$n_k = \bigwedge_{i=1}^m N(A_{ki}, A_{0i}) (k = 1, 2, \dots, n),$$

并有自然数 $k_0 (1 \leq k_0 \leq n)$ 使得 $n_{k_0} = \bigwedge_{k=1}^n n_k$, 则模式 A_0 隶属于 A_{k_0} .

最后值得我们注意的是模式识别与模糊聚类分析的关系和区别. 首先,二者都是研究模糊分类问题的方法,但二者既有关联,又有差别. 模糊聚类分析所研究的对象是一组样本,没事先确定的模式标准,只是根据对象的特性进行适当的分类. 而模糊模式识别所讨论的问题事先已知若干标准模式,或标准模式库,据此,对要待识别的对象进行识别,看它应属于哪一类. 因此,模糊聚类分析是一种无标准模式的分类方法,而模糊模式识别是一种有标准模式的分类方法.

另一方面,模糊聚类分析与模糊模式识别也是有关系的. 实际中,我们用模糊聚类分析法进行判别、预测的过程,事实上就是模糊聚类与模糊识别综合运用. 模糊识别中的标准模式就是在模糊聚类分析过程中得到的,即模糊聚类为模糊识别提供了标准模式库.

19.5 模糊综合评判方法

模糊综合评判是模糊决策中最常用的一种有效方法. 在实际中,常常需要一个事物做出评价(或评估),一般都涉及多个因素或多个指标,此时就要求我

们根据这些因素对事物做出综合评价,这就是所谓的综合评判,即综合评判就是要对受多个因素影响的事物(或对象)做出全面的评价,故模糊综合评判又称为模糊综合决策或模糊多元决策.传统的评判方法有总评分法和加权评分法.

总评分法:根据评判对象的评价项目 $u_i (i=1,2,\cdots,n)$,首先,对每个项目确定出评价的等级和相应的评分数 $s_i (i=1,2,\cdots,n)$,并将所有项目的分数求和 $S = \sum_{i=1}^n s_i$,然后,按总分的大小排序,从而确定出方案的优劣.

加权评分法:根据评判对象的诸多因素(或指标) $u_i (i=1,2,\cdots,n)$ 所处的地位或所起的作用一般不尽相同.因此,引入权重的概念,求其诸多因素(指标)评分 $s_i (i=1,2,\cdots,n)$ 的加权和 $S = \sum_{i=1}^n w_i s_i$. 其中 w_i 为第 $i (i=1,2,\cdots,n)$ 个因素(指标)的权值.

19.5.1 模糊综合评判的一般方法

1. 模糊综合评判的一般提法

设 $U = \{u_1, u_2, \cdots, u_n\}$ 为研究对象的 n 种因素(或指标),称之为**因素集**(或**指标集**). $V = \{v_1, v_2, \cdots, v_m\}$ 为诸因素(或指标)的 m 种评判所构成的评判集(或称评语集、评价集、决策集等),它们的元素个数和名称均可根据实际问题的需要和决策人主观确定.实际中,很多问题的因素评判集都是模糊的,因此,综合评判应该是 V 上的一个模糊子集

$$B = (b_1, b_2, \cdots, b_m) \in F(V),$$

其中 b_k 为评判 v_k 对模糊子集 B 的隶属度: $\mu_B(v_k) = b_k (k=1,2,\cdots,m)$,即反映了第 k 种评判 v_k 在综合评价中所起的作用.综合评判 B 依赖于各因素的权重,

即它应该是 U 上的模糊子集 $A = (a_1, a_2, \cdots, a_n) \in F(U)$, 且 $\sum_{i=1}^n a_i = 1$, 其中 a_i 表示第 i 种因素的权重.于是,当权重 A 给定以后,则相应地就可以给定一个综合评判 B .

2. 模糊综合评判的一般步骤

- (1) 确定因素集 $U = \{u_1, u_2, \cdots, u_n\}$;
- (2) 确定评判集 $V = \{v_1, v_2, \cdots, v_m\}$;
- (3) 确定模糊评判矩阵 $R = (r_{ij})_{n \times m}$:

首先,对每一个因素 u_i 做一个评判 $f(u_i) (i=1,2,\cdots,n)$,则可以得 U 到 V 的一个模糊映射 f ,即

$$f: U \rightarrow F(V), u_i \mapsto f(u_i) = (r_{i1}, r_{i2}, \cdots, r_{im}) \in F(V).$$

然后,由模糊映射 f 可以诱导出模糊关系 $R_f \in F(U \times V)$,即

$$R_f(u_i, v_j) = f(u_i)(v_j) = r_{ij} (i = 1, 2, \dots, n; j = 1, 2, \dots, m).$$

因此,可以确定出模糊评判矩阵 $R = (r_{ij})_{n \times m}$. 而且称 (U, V, R) 为模糊综合评判模型, U, V, R 称为该模型的三要素.

(4) 综合评判: 对于权重 $A = (a_1, a_2, \dots, a_n) \in F(U)$, 用模型 $M(\wedge, \vee)$ 取最大—最小合成运算, 可以得到综合评判

$$B = A \circ R \text{ (或 } b_j = \bigvee_{i=1}^n (a_i \wedge r_{ij}), j = 1, 2, \dots, m)$$

注: 关于评判集 V 的权重 $A = (a_1, a_2, \dots, a_n)$ 的确定在综合评判中起重要的作用, 通常情况下可以由决策人凭经验给出, 但往往带有一定的主观性. 要从实际出发, 或更客观地反映实际情况可采用专家评估法、加权统计法和频数统计法, 或更一般的模糊协调决策法、模糊关系方法等来确定.

19.5.2 综合评判模型的构成

如果模糊综合评判模型为 (U, V, R) , 对于权重 $A = (a_1, a_2, \dots, a_n) \in F(U)$, 模糊评判矩阵为 $R = (r_{ij})_{n \times m}$, 则用模型 $M(\wedge, \vee)$ 运算得综合评判为 $B = A \circ R =$

$$(b_1, b_2, \dots, b_m) \in F(V), \text{ 其中 } b_j = \bigvee_{i=1}^n (a_i \wedge r_{ij}) \quad (j = 1, 2, \dots, m).$$

事实上, 由于 $\sum_{i=1}^n a_i = 1$, 对于某些情况可能会出现 $a_i \leq r_{ij}$, 即 $a_i \wedge r_{ij} = a_i$. 这样可能导致模糊评判矩阵 R 中的许多信息的丢失, 即人们对某些因素 u_i 所作的评判信息在决策中未得到充分的利用. 从而导致综合评判结果失真. 为此, 实际中可以对模型 $M(\wedge, \vee)$ 进行改进.

(1) 模型 $M(\bullet, \vee)$ 法: 对于 $A = (a_1, a_2, \dots, a_n) \in F(U)$ 和 $R = (r_{ij})_{n \times m}$, 则用模型 $M(\bullet, \vee)$ 运算得 $B = A * R$, 即 $b_j = \bigvee_{i=1}^n (a_i \bullet r_{ij}) \quad (j = 1, 2, \dots, m).$

(2) 模型 $M(\wedge, +)$ 法: 对于 $A = (a_1, a_2, \dots, a_n) \in F(U)$ 和 $R = (r_{ij})_{n \times m}$, 则用模型 $M(\wedge, +)$ 运算得 $B = A * R$, 即 $b_j = \sum_{i=1}^n (a_i \wedge r_{ij}) \quad (j = 1, 2, \dots, m).$

(3) 模型 $M(\bullet, +)$ 法: 对于 $A = (a_1, a_2, \dots, a_n) \in F(U)$ 和 $R = (r_{ij})_{n \times m}$, 则用模型 $M(\bullet, +)$ 运算得 $B = A * R$, 即 $b_j = \sum_{i=1}^n (a_i \bullet r_{ij}) \quad (j = 1, 2, \dots, m).$

在实际应用时, 主因素 (即权重最大的因素) 在综合中起主导作用时, 则可选“主因素决定型”模型 $M(\wedge, \vee)$; 当模型 $M(\wedge, \vee)$ 失效时, 再来选用“主因素突出型”模型 $M(\bullet, \vee)$ 和 $M(\wedge, +)$; 当需要对所有因素的权重均衡时, 可选用加权平均模型 $M(\bullet, +)$. 在模型的选择时, 还要特别注意实际需求.

19.5.3 多层次模糊综合评判

对于实际中的许多问题往往都是涉及因素多,各因素的权重分配较为均衡的情况,此时,可采用将诸因素分为若干个层次进行研究.即首先分别对单层次各因素进行评判,然后再对所有的各层次因素作综合评判.这里仅就两个层次的情况进行说明,具体方法如下:

(1) 将因素集 $U = \{u_1, u_2, \dots, u_n\}$ 分成若干个组 $U_1, U_2, \dots, U_k (1 \leq k \leq n)$ 使得 $U = \bigcup_{i=1}^k U_i$, 且 $U_i \cap U_j = \emptyset (i \neq j)$, 称 $U = \{U_1, U_2, \dots, U_k\}$ 为一级因素集.

不妨设 $U_i = \{u_1^{(i)}, u_2^{(i)}, \dots, u_{n_i}^{(i)}\} (i = 1, 2, \dots, k; \sum_{i=1}^k n_i = n)$, 称之为二级因素集.

(2) 设评判集 $V = \{v_1, v_2, \dots, v_m\}$, 对二级因素集 $U_i = \{u_1^{(i)}, u_2^{(i)}, \dots, u_{n_i}^{(i)}\}$ 的 n_i 个因素进行单因素评判, 即建立模糊映射

$$f_i: U_i \rightarrow F(V), u_j^{(i)} \mapsto f_i(u_j^{(i)}) = (r_{j1}^{(i)}, r_{j2}^{(i)}, \dots, r_{jm}^{(i)}) (j = 1, 2, \dots, n_i),$$

于是得到评判矩阵为

$$R_i = \begin{bmatrix} r_{11}^{(i)} & r_{12}^{(i)} & \cdots & r_{1m}^{(i)} \\ r_{21}^{(i)} & r_{22}^{(i)} & \cdots & r_{2m}^{(i)} \\ \vdots & \vdots & & \vdots \\ r_{n_i 1}^{(i)} & r_{n_i 2}^{(i)} & \cdots & r_{n_i m}^{(i)} \end{bmatrix}.$$

不妨设 $U_i = \{u_1^{(i)}, u_2^{(i)}, \dots, u_{n_i}^{(i)}\}$ 的权重为 $A_i = (a_1^{(i)}, a_2^{(i)}, \dots, a_{n_i}^{(i)})$, 则可以求得综合评判为

$$B_i = A_i \circ R_i = (b_1^{(i)}, b_2^{(i)}, \dots, b_m^{(i)}) (i = 1, 2, \dots, k),$$

其中 $b_j^{(i)}$ 由模型 $M(\wedge, \vee)$, 或 $M(\bullet, \vee)$ 、 $M(\wedge, +)$ 、 $M(\bullet, +)$ 确定.

(3) 对于一级因素集 $U = \{U_1, U_2, \dots, U_k\}$ 作综合评判, 不妨设其权重 $A = (a_1, a_2, \dots, a_k)$, 总评判矩阵为 $R = [B_1, B_2, \dots, B_k]^T$. 按模型 $M(\wedge, \vee)$ 或 $M(\bullet, \vee)$ 、 $M(\wedge, +)$ 、 $M(\bullet, +)$ 运算得到综合评判 $B = A \circ R = (b_1, b_2, \dots, b_m) \in F(V)$.

19.6 中介服务机构的信誉评估问题^[10]

19.6.1 问题的提出

近年来,社会上出现了很多不同规模、不同性质的中介服务机构,但目前还缺少规范统一管理的政策和法规,有些地方由此产生了许多社会问题. 为了加强对这些机构的管理,政府有关部门需要对这些机构的“信誉”做出客观地评价,

以便制定相应的政策和法规。

据调查,影响“信誉”的因素有(1)遵纪守法情况、(2)纳税情况、(3)奖惩情况、(4)治安情况、(5)履行合同情况、(6)中介成功率、(7)纠纷情况、(8)服务质量、(9)客户意见、(10)社会效益、(11)机构影响、(12)公益事业、(13)内部管理、(14)经济情况、(15)服务项目等。

现在的问题是依据这些定量或定性因素的具体含义,建立对中介服务机构“信誉”的评估模型,并给政府主管部门提出实施建议。

19.6.2 问题的分析与假设

首先,我们注意到“信誉”本身就是一个模糊的概念,与之相关的15个因素有些是定量的,有些是定性的,对“信誉”影响的程度不尽相同,而且,具有不同的实际含义。为此,我们可以用综合模糊评判方法对中介服务机构的信誉给出一种综合评估方案。

考虑到各因素之间的关系,按因素的不同性质分为四个因素集(即主要因素):

机构的法纪情况: $A = \{(1), (2), (3), (4)\}$;

机构的业务情况: $B = \{(5), (6), (7), (8), (9)\}$;

机构的社会效应: $C = \{(10), (11), (12)\}$;

机构的内部情况: $D = \{(13), (14), (15)\}$ 。

根据问题的实际情况,对四个因素集中的15个二级因素,又可以分为31个三级因素(如表19-2),对每个三级因素都可给出一个评判集,相应地可以构造出二级因素的模糊矩阵和各因素的权向量,确定出各主要因素的隶属度,从而可求出服务机构的信誉度,对服务机构的信誉给出客观的评价。

表 19-2 各级因素及其权值

| 主要因素 | 二级因素 | 权 重 | 模糊矩阵 | 三 级 因 素 | 权 重 |
|------------------------------------|---------------------------------|-------------|-----------|--------------------------------|----------------|
| (A) 法 纪 情 况 0.3 | (A ₁) 遵纪 守法情况 | $a_1 = 0.3$ | R_{A_1} | (A ₁₁) 经营活动 | $a_{11} = 0.5$ |
| | | | | (A ₁₂) 财务制度 | $a_{12} = 0.5$ |
| | (A ₂) 纳 税 情 况 | $a_2 = 0.5$ | R_{A_2} | (A ₂₁) 所纳税与应纳税的比率 | $a_{21} = 0.6$ |
| | | | | (A ₂₂) 逃税次数 | $a_{22} = 0.2$ |
| | | | | (A ₂₃) 逃税罚金与所纳税的比率 | $a_{23} = 0.2$ |
| | (A ₃) 奖惩 情 况 | $a_3 = 0.1$ | R_{A_3} | (A ₃₁) 奖励的次数 | $a_{31} = 0.5$ |
| | | | | (A ₃₂) 惩罚的次数 | $a_{32} = 0.5$ |
| | (A ₄) 治安 情 况 | $a_4 = 0.1$ | R_{A_4} | (A ₄₁) 发生治安案件的次数 | $a_{41} = 0.4$ |
| | | | | (A ₄₂) 发生刑事案件的次数 | $a_{42} = 0.6$ |

续表

| 主要因素 | 二级因素 | 权 重 | 模糊矩阵 | 三 级 因 素 | 权 重 |
|--------------------|----------------|-------------|-----------|-----------------------------------|----------------|
| (B) 业务情况 0.3 | (B_1) 履行合同情况 | $b_1 = 0.2$ | R_{B_1} | (B_{11}) 履行合同与全部合同比率 | $b_{11} = 0.6$ |
| | | | | (B_{12}) 主动违约占合同的比率 | $b_{12} = 0.4$ |
| | (B_2) 纳税情况 | $b_2 = 0.2$ | R_{B_2} | (B_{21}) 中介服务的成功率 | $b_{21} = 0.6$ |
| | | | | (B_{22}) 因机构原因不成功事件的占全部不成功事件的比率 | $b_{22} = 0.4$ |
| | (B_3) 奖惩情况 | $b_3 = 0.2$ | R_{B_3} | (B_{31}) 发生纠纷的次数 | $b_{31} = 0.6$ |
| | | | | (B_{32}) 调解成功的比率 | $b_{32} = 0.4$ |
| | (B_4) 服务质量 | $b_4 = 0.2$ | R_{B_4} | (B_{41}) 客户利益 | $b_{41} = 0.6$ |
| | | | | (B_{42}) 后续服务 | $b_{42} = 0.4$ |
| | (B_5) 客户意见 | $b_5 = 0.2$ | R_{B_5} | (B_{51}) 客户表扬的次数 | $b_{51} = 0.5$ |
| | | | | (B_{52}) 客户投诉的次数 | $b_{52} = 0.5$ |
| (C) 社会效益 0.2 | (C_1) 社会效益 | $c_1 = 0.5$ | R_{C_1} | (C_{11}) 对社会的进步和发展所产生的影响 | $c_{11} = 1$ |
| | (C_2) 机构影响 | $c_2 = 0.3$ | R_{C_2} | (C_{21}) 服务项目的覆盖范围 | $c_{21} = 0.5$ |
| | | | | (C_{22}) 公共关系 | $c_{22} = 0.5$ |
| | (C_3) 公益事业 | $c_3 = 0.2$ | R_{C_3} | (C_{31}) 参加公益事业活动的次数 | $c_{31} = 0.4$ |
| | | | | (C_{32}) 公益事业费与总收入比率 | $c_{32} = 0.6$ |
| (D) 内部情况 0.2 | (D_1) 内部管理 | $d_1 = 0.5$ | R_{D_1} | (D_{11}) 领导的管理水平 | $d_{11} = 0.4$ |
| | | | | (D_{12}) 职员的业务水平 | $d_{12} = 0.4$ |
| | | | | (D_{13}) 人员的凝聚力 | $d_{13} = 0.2$ |
| | (D_2) 经济情况 | $d_2 = 0.3$ | R_{D_2} | (D_{21}) 收入情况 | $d_{21} = 0.4$ |
| | | | | (D_{22}) 还贷情况 | $d_{22} = 0.3$ |
| | | | | (D_{23}) 拖欠情况 | $d_{23} = 0.3$ |
| | (D_3) 服务项目 | $d_3 = 0.2$ | R_{D_3} | (D_{31}) 中介服务的项目数 | $d_{31} = 1$ |

19.6.3 模型的建立与求解

因为影响“信誉”的有 A 、 B 、 C 、 D 四个主要因素, 15 个二级因素, 31 个三级因素, 根据实际可以给出各因素的权重(如表 19-2). 对于每个三级因素都可给出由 5 个元素组成的评判集, 例如: {很好, 好, 一般, 差, 很差}、{很可靠, 可靠, 合法, 基本合法, 非法}、{很高, 高, 一般, 低, 很低}、{很多, 多, 一般, 少, 很少}、

{100% , [95% , 100%] , [90% , 95%] , [80% , 90%] , <80% } 等(如表 19-3)。

表 19-3 各因素的评判集

| 因 素 | 等 级 1 | 等 级 2 | 等 级 3 | 等 级 4 | 等 级 5 |
|----------------------------------|-------|---------------|--------------|--------------|-------|
| A_{11} A_{12} | 很可靠 | 可靠 | 合法 | 基本合法 | 非法 |
| A_{21} | 100% | [95% , 100%) | [90% , 95%) | [80% , 90%) | <80% |
| A_{22} | 0 | 1 | 2 | 3 | >3 |
| A_{23} | 0 | (0,2%] | (2% ,5%] | (5% ,10%] | >10% |
| A_{31} | 5 | 4 | 2 | 1 | 0 |
| A_{32} | 0 | 1 | 2 | 4 | >4 |
| A_{41} | 0 | 0 | 1 | 3 | >3 |
| A_{42} | 0 | 0 | 1 | 2 | >2 |
| B_{11} B_{12} | 100% | [95% , 100%) | [90% , 95%) | [80% , 90%) | <80% |
| | 0 | (0,1%] | (1% ,5%] | (5% ,10%] | >10% |
| B_{21} | 100% | [90% , 100%) | [70% , 90%) | [50% , 70%) | <50% |
| B_{22} | 0 | (0,2%] | (2% ,5%] | (5% ,10%] | >10% |
| B_{31} | 0 | 1 | 2 | 5 | >5 |
| B_{32} | 100% | [95% , 100%) | [90% , 95%) | [80% , 90%) | <80% |
| B_{41} | 很高 | 高 | 一般 | 低 | 很低 |
| B_{42} | 很好 | 好 | 一般 | 差 | 很差 |
| B_{51} | >10 | (5,10] | (2,5] | (0,2] | 0 |
| B_{52} | 0 | 1 | 2 | 3 | >3 |
| C_{11} | 很大 | 大 | 一般 | 小 | 很小 |
| C_{21} | 很广 | 广 | 一般 | 窄 | 很窄 |
| C_{22} | 很好 | 好 | 一般 | 差 | 很差 |
| C_{31} | >10 | (5,10] | (2,5] | (0,2] | 0 |
| C_{32} | >5% | (1% ,5%] | (0.5% ,1%] | (0,0.5%] | 0 |
| D_{11} D_{12} D_{13} | 很高 | 高 | 一般 | 低 | 很低 |
| D_{21} | 很高 | 高 | 一般 | 低 | 很低 |
| D_{22} | 100% | [95% , 100%) | [90% , 95%) | [80% , 90%) | <80% |
| D_{23} | 0 | 很少 | 少 | 多 | 很多 |
| D_{31} | 很多 | 多 | 一般 | 少 | 很少 |

实际中,通常是由专家组根据评判集对服务机构进行评判,由评判结果构造出每个因素的评判和模糊矩阵.例如:记第 i 个三级因素的评判为 $(r_{i1}, r_{i2}, r_{i3}, r_{i4}, r_{i5})$ ($i=1, 2, \dots, m$),则相应二级因素的模糊评判矩阵为

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{15} \\ r_{21} & r_{22} & \cdots & r_{25} \\ \vdots & \vdots & & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{m5} \end{bmatrix},$$

其中

$$r_{ij} = \frac{\text{评其为第 } j \text{ 个等级的人数}}{\text{评委的总人数}} \begin{pmatrix} i = 1, 2, \dots, m; \\ j = 1, 2, \dots, 5 \end{pmatrix}$$

为第 i 个三级因素第 j 个等级的隶属度.于是可得15个二级因素的模糊评判矩阵 $R_{A_i}, R_{B_j}, R_{C_k}, R_{D_l}$,则相应的评语向量分别为

$$V_A^{(i)} = W_{A_i} \cdot R_{A_i}, V_B^{(j)} = W_{B_j} \cdot R_{B_j}, V_C^{(k)} = W_{C_k} \cdot R_{C_k}, V_D^{(l)} = W_{D_l} \cdot R_{D_l}$$

其中, $i=1, 2, 3, 4; j=1, 2, \dots, 5; k=1, 2, 3; l=1, 2, 3$.故四个主要因素的模糊评判矩阵为

$$R_A = (V_A^{(1)}, V_A^{(2)}, V_A^{(3)}, V_A^{(4)})^T, R_B = (V_B^{(1)}, V_B^{(2)}, V_B^{(3)}, V_B^{(4)}, V_B^{(5)})^T$$

$$R_C = (V_C^{(1)}, V_C^{(2)}, V_C^{(3)})^T, R_D = (V_D^{(1)}, V_D^{(2)}, V_D^{(3)})^T$$

则相应的评语向量为

$$V_A = (a_1, a_2, a_3, a_4) \cdot R_A = (v_{a_1}, v_{a_2}, v_{a_3}, v_{a_4}, v_{a_5}),$$

$$V_B = (b_1, b_2, b_3, b_4, b_5) \cdot R_B = (v_{b_1}, v_{b_2}, v_{b_3}, v_{b_4}, v_{b_5}),$$

$$V_C = (c_1, c_2, c_3) \cdot R_C = (v_{c_1}, v_{c_2}, v_{c_3}, v_{c_4}, v_{c_5}),$$

$$V_D = (d_1, d_2, d_3) \cdot R_D = (v_{d_1}, v_{d_2}, v_{d_3}, v_{d_4}, v_{d_5}).$$

令 $\mu_1 = \sum_{j=1}^5 v_{a_j}, \mu_2 = \sum_{j=1}^5 v_{b_j}, \mu_3 = \sum_{j=1}^5 v_{c_j}, \mu_4 = \sum_{j=1}^5 v_{d_j}$,则 $\mu_1, \mu_2, \mu_3, \mu_4$ 分别是

主要因素为 A, B, C, D 的隶属度,其评价向量记为 $\mu = (\mu_1, \mu_2, \mu_3, \mu_4)^T$.于是可得对中介服务机构“信誉”的综合评价指标为

$$D = W \cdot \mu = 0.3\mu_1 + 0.3\mu_2 + 0.2\mu_3 + 0.2\mu_4.$$

19.6.4 模型的检验与分析

我们对该模型在计算机上分别以 8 名和 12 名专家组成的评委会随机地对 50 个中介服务机构进行了评判检验,结果表明都能客观地给出综合评价指标,为此,这一模型是可以应用于实际的,而且具有相当广泛的通用性。

在模型中关于各因素的权重主要是依据实际背景主观上给定的,可能带有片面性,实际中可以根据侧重点适当调整,不影响整体效果。如果能详细地分析各因素之间的关系,利用 AHP 方法通过两两比较矩阵确定权向量,结果可能会更客观,但工作量会增大。

关于各因素隶属度的确定,采用专家组评判的方法是切合实际的,当然,实际操作中专家组成员可以由各相关方面的领导、专家或代表参加,应具有权威性和客观性。另外,关于评判集的确定也不是惟一的。

19.7 参考案例与参考文献

1. 参考案例

- (1) 模糊聚类分析在市场划分中的应用——文献[1]:114—1118
- (2) 模糊聚类分析人才选拔中的应用——文献[6]:644—650
- (3) 大学生体质水平的模糊识别问题——文献[1]:132—135
- (4) 小麦亲本的模糊识别问题——文献[1]:159—163
- (5) 蔬菜品质的模糊综合评判问题——文献[1]:240—249
- (6) 导弹效能分析的模糊综合评判问题——文献[7]:426—430

2. 参考文献

- [1] 谢季坚,刘承平. 模糊数学方法及其应用. 武汉:华中科技大学出版社,2000
- [2] 彭祖赠,孙韞玉. 模糊(Fuzzy)数学及其应用. 武汉:武汉大学出版社,2002
- [3] 杨纶标,高英仪. 模糊数学原理及其应用. 广州:华南理工大学出版社,2003
- [4] 刘普寅,吴孟达. 模糊理论及其应用. 长沙:国防科技大学出版社,1998
- [5] 刘增良. 模糊技术与应用选编(1). 北京:北京航空航天大学出版社,1997
- [6] 刘增良. 模糊技术与应用选编(2). 北京:北京航空航天大学出版社,1997
- [7] 刘增良. 模糊技术与应用选编(3). 北京:北京航空航天大学出版社

社,1998

[8] 刘增良. 模糊技术与神经网络技术应用选编(4). 北京:北京航空航天大学出版社,1999

[9] 刘增良. 模糊技术与神经网络技术应用选编(5). 北京:北京航空航天大学出版社,2001

[10] 韩中庚,张彦肖. 中介服务机构的信誉评估模型. 数学的实践与认识, 2001,5(31):517—518

第二十章 灰色系统分析方法

我们通常所说的系统是指:由客观世界中相同或相似的事物和因素按一定的秩序相互关联、相互制约而构成的一个整体.例如:工程技术系统、社会系统、经济系统等等.如果一个系统中具有充足的信息量,其发展变化的规律明显、定量描述方便、结构与参数具体,则这种系统通常称为白色系统.如果一个系统的内部特性全部是未知的,则称此系统为黑色系统.介于白色系统和黑色系统之间的称为灰色系统(Grey System),即系统内部信息和特性是部分已知的,另一部分是未知的.例如:像社会系统、农业系统、经济系统、气象系统、生物系统等.对于这类系统,人们往往难以建立客观的物理原形,内部因素难以辨识,以及相互之间的关系较为隐蔽,人们也难已准确了解这类系统的行为特征.因此,对于这类问题进行定量描述,即建立模型难度较大.区别白色系统与灰色系统的重要标志是系统内各因素之间是否具有确定的关系.

灰色系统分析方法主要是根据具体灰色系统的行为特征数据,充分利用数量不多的数据和信息寻求相关因素自身与各因素之间的数学关系,即建立相应的数学模型.目前,灰色系统理论在实际中已得到了广泛的应用,例如:在工程技术、经济管理、气象预报,以及政治、社会、工业、农业等领域都取得了一定应用成果.

20.1 灰色系统分析的基本概念

20.1.1 灰数的概念及其运算

1. 灰数及其表示法

灰数是灰色系统的标志之一.所谓的灰数是指信息不完全的数,例如:“那个小姑娘的身高大约有 165 cm 左右,体重只有 40 kg 左右”.这里的 165 cm 左右和 40 kg 左右都是灰数,可以分别记为 $\otimes(165)$ 和 $\otimes(40)$.再如:“他的体温大约在 38 度~39 度之间”,关于体温是灰数,记为 $\otimes(T) \in [38, 39]$.

如果 $\widetilde{\otimes}$ 为灰数 \otimes 的白化默认(即对形象、形态、实体、数字的默认)数,简称

为白化数,则灰数 \otimes 是白化数 $\widetilde{\otimes}$ 的全体.

如果 \otimes 是离散灰数,则有

$$\forall \widetilde{\otimes} \in \otimes \Rightarrow \widetilde{\otimes} \in A = \{x(k) \mid k \in K = \{1, 2, \dots, n\}\}.$$

如果灰数 \otimes 中的白化数是按区间连续分布的,则有

$$\forall \widetilde{\otimes} \in \otimes \Rightarrow \widetilde{\otimes} \in It(a, b) \in \{[a, b], (a, b), [a, b), (a, b]\}.$$

2. 灰数的运算

(1) 离散灰数运算:

设 \otimes_i 与 \otimes_j 为两个离散的灰数,即

$$\forall \widetilde{\otimes}_i \in \otimes_i \Rightarrow \widetilde{\otimes}_i \in \{x_i(k) \mid k \in K = \{1, 2, \dots, n\}\},$$

$$\forall \widetilde{\otimes}_j \in \otimes_j \Rightarrow \widetilde{\otimes}_j \in \{x_j(l) \mid l \in L = \{1, 2, \dots, m\}\},$$

则

1) \otimes_i 与 \otimes_j 的加运算为

$$\begin{aligned} \forall \widetilde{\otimes}_i + \widetilde{\otimes}_j \in \otimes_i + \otimes_j \Rightarrow \widetilde{\otimes}_i + \widetilde{\otimes}_j \in \{x_i(k) + x_j(l) \mid k \in K, l \in L\} \\ \subset [\min_k \min_l \{x_i(k) + x_j(l)\}, \max_k \max_l \{x_i(k) + x_j(l)\}]; \end{aligned}$$

2) \otimes_i 与 \otimes_j 的减运算为

$$\begin{aligned} \forall \widetilde{\otimes}_i - \widetilde{\otimes}_j \in \otimes_i - \otimes_j \Rightarrow \widetilde{\otimes}_i - \widetilde{\otimes}_j \in \{x_i(k) - x_j(l) \mid k \in K, l \in L\} \\ \subset [\min_k \min_l \{x_i(k) - x_j(l)\}, \max_k \max_l \{x_i(k) - x_j(l)\}]; \end{aligned}$$

3) \otimes_i 与 \otimes_j 的乘运算为

$$\begin{aligned} \forall \widetilde{\otimes}_i \times \widetilde{\otimes}_j \in \otimes_i \times \otimes_j \Rightarrow \widetilde{\otimes}_i \times \widetilde{\otimes}_j \in \{x_i(k) \cdot x_j(l) \mid k \in K, l \in L\} \\ \subset [\min_k \min_l \{x_i(k) \cdot x_j(l)\}, \max_k \max_l \{x_i(k) \cdot x_j(l)\}]; \end{aligned}$$

4) \otimes_i 与 \otimes_j 的除运算为

$$\begin{aligned} \forall \widetilde{\otimes}_i / \widetilde{\otimes}_j \in \otimes_i / \otimes_j \Rightarrow \widetilde{\otimes}_i / \widetilde{\otimes}_j \in \{x_i(k) / x_j(l) \mid k \in K, l \in L\} \\ \subset [\min_k \min_l \{x_i(k) / x_j(l)\}, \max_k \max_l \{x_i(k) / x_j(l)\}]. \end{aligned}$$

(2) 连续灰数的运算:

设 \otimes_i 与 \otimes_j 为两个连续的灰数,即

$$\forall \widetilde{\otimes}_i \in \otimes_i \Rightarrow \widetilde{\otimes}_i \in It(a_i, b_i),$$

$$\forall \widetilde{\otimes}_j \in \otimes_j \Rightarrow \widetilde{\otimes}_j \in It(a_j, b_j),$$

则

1) \otimes_i 与 \otimes_j 的加运算为

$$\forall \widetilde{\otimes}_i + \widetilde{\otimes}_j \in \otimes_i + \otimes_j \Rightarrow \widetilde{\otimes}_i + \widetilde{\otimes}_j \in It(a_i + a_j, b_i + b_j);$$

2) \otimes_i 与 \otimes_j 的减运算为

$$\forall \widetilde{\otimes}_i - \widetilde{\otimes}_j \in \otimes_i - \otimes_j \Rightarrow \widetilde{\otimes}_i - \widetilde{\otimes}_j \in It(\min\{a_i - b_j, b_i - a_j\}, \max\{a_i - b_j, b_i - a_j\});$$

3) \otimes_i 与 \otimes_j 的乘运算为

$$\begin{aligned} \forall \widetilde{\otimes}_i \times \widetilde{\otimes}_j &\in \otimes_i \times \otimes_j \\ \Rightarrow \widetilde{\otimes}_i \times \widetilde{\otimes}_j &\in It(\min\{a_i a_j, a_i b_j, a_j b_i, b_i b_j\}, \\ &\max\{a_i a_j, a_i b_j, a_j b_i, b_i b_j\}); \end{aligned}$$

4) \otimes_i 与 \otimes_j 的除运算为

$$\begin{aligned} \forall \widetilde{\otimes}_i / \widetilde{\otimes}_j &\in \otimes_i / \otimes_j \\ \Rightarrow \widetilde{\otimes}_i / \widetilde{\otimes}_j &\in It\left(\min\left\{\frac{a_i}{a_j}, \frac{a_i}{b_j}, \frac{b_i}{a_j}, \frac{b_i}{b_j}\right\}, \right. \\ &\left. \max\left\{\frac{a_i}{a_j}, \frac{a_i}{b_j}, \frac{b_i}{a_j}, \frac{b_i}{b_j}\right\}\right). \end{aligned}$$

上述某些运算也可以推广到任意有限个灰数的情况,另外,作为特殊情况,类似地可以给出数与灰数的混合运算^[1].

20.1.2 灰色关联分析

实际中,经常需要对系统相关的因素进行分析研究,即分析相关各因素对系统哪些是主要的,哪些是次要的;哪些是直接的,哪些是间接的等等.统计分析中的相关分析等方法是分析研究各因素之间关联程度的一种有效方法,但它往往是需要有大量的统计数据,计算量大,而且可能会出现反常的情况.为此,针对灰色系统采用关联度分析的方法来研究相应的问题.

1. 单因子的情况

如果系统的行为只有一个因子 x_0 ,而 x_0 受到多种因素 $x_i (i=1,2,\dots,n)$ 的影响,一种利用因素 x_i 对因子 x_0 的灰关联度来表示 x_i 对 x_0 影响大小的方法,则称为灰关联分析.

设系统行为因子 x_0 的参考数列为

$$x_0 = \{x_0(k) \mid k = 1, 2, \dots, n\} = (x_0(1), x_0(2), \dots, x_0(n)),$$

相关因素为 $x_i (i=1, 2, \dots, n)$, 即比较数列为

$$x_i = \{x_i(k) \mid k = 1, 2, \dots, n\} = (x_i(1), x_i(2), \dots, x_i(n)) \quad (i = 1, 2, \dots, m),$$

则参考数列对于各比较数列间的绝对差为

$$\Delta_i(k) = |x_0(k) - x_i(k)| \quad (k = 1, 2, \dots, n; 1 \leq i \leq m),$$

记 $\Delta_i = (\Delta_i(1), \Delta_i(2), \dots, \Delta_i(n))$, 称之为差数列. 于是, 可以定义比较数列 x_i

对参考数列 x_0 在第 k 点的灰关联系数为

$$r(x_0(k), x_i(k)) = \frac{\min_i \min_k \Delta_i(k) + \alpha \max_i \max_k \Delta_i(k)}{\Delta_i(k) + \alpha \max_i \max_k \Delta_i(k)}, \quad (20.1)$$

其中常数 $\alpha \in [0, 1]$, 称为分辨率系数. 显然, 当 α 越大时, 分辨率越大; 当 α 越小时, 分辨率越小, 一般情况取 $\alpha = 0.5$.

对于所有的点 $k = 1, 2, \dots, n$, 则定义比较数列 x_i 对参考数列 x_0 的灰关联度为

$$r_i = r(x_0, x_i) = \frac{1}{n} \sum_{k=1}^n r(x_0(k), x_i(k)) \quad (i = 1, 2, \dots, m), \quad (20.2)$$

即用灰关联度 r_i 可以表示因素 x_i 对行为因子 x_0 的关联(影响)程度.

2. 多因子的情况

设系统行为有多个因子, 不妨设因子集为 $X = \{x_i | i = 1, 2, \dots, l\}$. 如果因素数列 x_i 满足下列条件, 则称 X 为灰关联因子集:

(1) 数列 x_i 的数据 $x_i(k)$ 之间具有数值可比性, 即指定 $x_i(k)$ 与 $x_j(t)$ 之间的数值是可以比较的, 或相等、或接近、或同数量级等;

(2) 数列 x_i 之间具有可接近性, 即非平等性;

(3) 数列 x_i 之间具有同级性, 即同为正(极大值)极性, 或负(极小值)极性, 或适中极性.

以灰关联因子集 X 中的一个因子 $x_i (1 \leq i \leq l)$ 为参考数列, 以任意因子 $x_j \in X$ 为比较数列, 则二者间的绝对差为

$$\Delta_{ij}(k) = |x_i(k) - x_j(k)| \quad (k = 1, 2, \dots, n; j = 1, 2, \dots, l),$$

相应的差数列为 $\Delta_{ij} = (\Delta_{ij}(1), \Delta_{ij}(2), \dots, \Delta_{ij}(n))$, 其比较数列 x_j 对参考数列 x_i 在第 k 点的灰关联为

$$r(x_i(k), x_j(k)) = \frac{\min_i \min_j \min_k \Delta_{ij}(k) + \alpha \max_i \max_j \max_k \Delta_{ij}(k)}{\Delta_{ij}(k) + \alpha \max_i \max_j \max_k \Delta_{ij}(k)},$$

其中常数 $\alpha \in [0, 1]$ 为分辨率系数. 于是有 x_j 对 x_i 的灰关联度为

$$r_{ij} = r(x_i, x_j) = \frac{1}{n} \sum_{k=1}^n r(x_i(k), x_j(k)) \quad (i = 1, 2, \dots, m; j = 1, 2, \dots, l).$$

20.1.3 灰色生成数列

在灰色系统理论中, 把一切随机量都看作灰色数, 即是在指定范围变化的所有白色数的全体. 对灰数的处理主要是利用数据处理方法去寻求数据间的内在规律, 通过对已知数据列中的数据进行处理而产生新的数据列, 以此来研究寻找

数据的规律性,这种方法称为**数据的生成**.数据的生成方式有多种,常用的方法有累加生成、累减生成和加权累加生成等.

1. 累加生成

把数列各时刻数据依次累加的过程称为**累加生成过程**,记为 AGO (Accumulated Generating Operation). 由累加生成过程所得到的新数列称为**累加生成数列**.

设原始数列为 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$, 令

$$x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i) \quad (k = 1, 2, \dots, n), \quad (20.3)$$

则称 $x^{(1)}(k)$ 为数列 $x^{(0)}$ 的**1次累加生成**, 数列 $x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$ 称为数列 $x^{(0)}$ 的**1次累加生成数列**. 类似地有

$$x^{(r)}(k) = \sum_{i=1}^k x^{(r-1)}(i) \quad (k = 1, 2, \dots, n, r \geq 1),$$

称之为 $x^{(0)}$ 的 **r 次累加生成**. 记 $x^{(r)} = (x^{(r)}(1), x^{(r)}(2), \dots, x^{(r)}(n))$, 称之为 $x^{(0)}$ 的 **r 次累加生成数列**.

2. 累减生成

对于原始数据列依次做前后相邻的两个数据相减的运算过程称为**累减生成过程**, 记为 IAGO. 如果原始数据列为 $x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$, 令

$$x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1), k = 2, 3, \dots, n,$$

则称 $x^{(0)}(k)$ 为数列 $x^{(1)}$ 的**1次累减生成**.

一般地, 对于 r 次累加生成数列 $x^{(r)} = (x^{(r)}(1), x^{(r)}(2), \dots, x^{(r)}(n))$ ($r \geq 1$), 则称

$$x^{(r-1)}(k) = x^{(r)}(k) - x^{(r)}(k-1), \quad k = 2, 3, \dots, n$$

为数列 $x^{(r)}$ 的 **r 次累减生成**.

3. 均值生成

设原始数列 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(k-1), x^{(0)}(k), \dots, x^{(0)}(n))$, 则称 $x^{(0)}(k-1)$ 与 $x^{(0)}(k)$ 为数列 $x^{(0)}$ 的**邻值**, $x^{(0)}(k-1)$ 为**后邻值**, $x^{(0)}(k)$ 为**前邻值**.

对于常数 $\alpha \in [0, 1]$, 则称

$$z^{(0)}(k) = \alpha x^{(0)}(k) + (1 - \alpha) x^{(0)}(k-1)$$

为由数列 $x^{(0)}$ 的邻值在生成系数(权) α 下的**邻值生成数(或生成值)**.

特别地, 当生成系数 $\alpha = 0.5$ 时, 则称

$$z^{(0)}(k) = 0.5 x^{(0)}(k) + 0.5 x^{(0)}(k-1) \quad (20.4)$$

为**邻均值生成数**, 即**等权邻值生成数**.

类似地, 可以定义**非邻值生成数**:

$$z^{(0)}(k) = \alpha x^{(0)}(k+1) + (1-\alpha)x^{(0)}(k-1)$$

和

$$z^{(0)}(k) = 0.5x^{(0)}(k+1) + 0.5x^{(0)}(k-1),$$

而数列 $z^{(0)} = (z^{(0)}(1), z^{(0)}(2), \dots, z^{(0)}(n))$ 称为均值 (mean) 生成数列.

20.2 灰色模型 GM

灰色系统理论是基于关联空间、光滑离散函数等概念定义灰导数与灰微分方程,进而用离散数据列建立微分方程形式的动态模型,由于这是本征灰色系统的基本模型,而且模型是近似的、非惟一的,故这种模型为灰色模型,记为 GM (Grey Model). 即灰色模型是利用离散随机数经过生成变为随机性被显著削弱而且较有规律的生成数,建立起的微分方程形式的模型,这样便于对其变化过程进行研究和描述.

20.2.1 GM(1,1)模型

1. GM(1,1)的定义

设 $x^{(0)}$ 为 n 个元素的数列 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$, $x^{(0)}$ 的 AGO 生成数列为 $x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$, 其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$ ($k = 1, 2, \dots, n$). 则定义 $x^{(1)}$ 的灰导数为

$$d(k) = x^{(0)}(k) = x^{(1)}(k) - x^{(1)}(k-1).$$

令 $z^{(1)}$ 为数列 $x^{(1)}$ 的均值 (mean) 数列, 即

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1) \quad (k = 2, 3, \dots, n),$$

则 $z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n))$. 于是定义 GM(1,1) 的灰微分方程模型为

$$d(k) + az^{(1)}(k) = b,$$

即

$$x^{(0)}(k) + az^{(1)}(k) = b, \quad (20.5)$$

其中 $x^{(0)}(k)$ 称为灰导数, a 称为发展系数, $z^{(1)}(k)$ 称为白化背景值, b 称为灰作用量.

将时刻 $k = 2, 3, \dots, n$ 代入 (20.5) 式中有,

$$\begin{cases} x^{(0)}(2) + az^{(1)}(2) = b, \\ x^{(0)}(3) + az^{(1)}(3) = b, \\ \dots\dots\dots \\ x^{(0)}(n) + az^{(1)}(n) = b. \end{cases}$$

令 $Y_N = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n))^T$, $u = (a, b)^T$, $B =$

$$\begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix},$$

称 Y_N 为数据向量, B 为数据矩阵, u 为参数向量, 则

GM(1,1) 可以表示为矩阵方程 $Y_N = B \cdot u$.

对于参数向量 u 的确定方法:

事实上, 如果存在 $(B^T \cdot B)^{-1}$, 由最小二乘法则有

$$\hat{u} = (\hat{a}, \hat{b})^T = (B^T \cdot B)^{-1} B^T Y_N.$$

具体地

$$\hat{a} = \frac{CD - (n-1)E}{(n-1)F - C^2}, \hat{b} = \frac{DF - CE}{(n-1)F - C^2},$$

其中 $C = \sum_{k=2}^n z^{(1)}(k)$, $D = \sum_{k=2}^n x^{(0)}(k)$, $E = \sum_{k=2}^n z^{(1)}(k)x^{(0)}(k)$, $F = \sum_{k=2}^n (z^{(1)}(k))^2$.

2. GM(1,1) 的白化型

对于 GM(1,1) 的灰微分方程 (20.5), 如果将 $x^{(0)}(k)$ 的时刻 $k=2, 3, \dots, n$ 视为连续的变量 t , 则数列 $x^{(1)}$ 就可以视为时间 t 的函数, 记为 $x^{(1)} = x^{(1)}(t)$, 并让灰导数 $x^{(0)}(k)$ 对应于导数 $\frac{dx^{(1)}}{dt}$, 背景值 $z^{(1)}(k)$ 对应于 $x^{(1)}(t)$. 于是得到

GM(1,1) 的灰微分方程对应的白微分方程为

$$\frac{dx^{(1)}}{dt} + ax^{(1)}(t) = b, \quad (20.6)$$

称之为 GM(1,1) 的白化型.

值得注意的是: GM(1,1) 的白化型 (20.6) 并不是由 GM(1,1) 的灰微分方程直接推导出来的, 它仅仅是一种“借用”、或“白化默认”. 所以 GM(1,1) 的白化型本身, 以及一切从白化型推导出来的结果, 只是在不与定义有矛盾的情况下成立, 否则无意义.

另一方面, GM(1,1) 的白化型是一个真正的微分方程, 如果白化型模型精度高, 则表明所用数列建立的模型 GM(1,1) 与真正的微分方程模型吻合较好, 反之亦然.

20.2.2 GM(1,N) 模型

1. GM(1,N) 模型定义

GM(1,1) 即表示模型是 1 阶的, 且只含 1 个变量的灰色模型. 而 GM(1,N)

即表示模型是 1 阶的, 包含有 N 个变量的灰色模型.

设系统有 N 个行为因子, 即原始数列为

$$x_i^{(0)} = (x_i^{(0)}(1), x_i^{(0)}(2), \dots, x_i^{(0)}(n)), i = 1, 2, \dots, N$$

记 $x_i^{(1)}$ 为 $x_i^{(0)}$ 的 AGO 数列, 即

$$\begin{aligned} x_i^{(1)} &= (x_i^{(1)}(1), x_i^{(1)}(2), \dots, x_i^{(1)}(n)) \\ &= (x_i^{(1)}(1), x_i^{(1)}(1) + x_i^{(0)}(2), \dots, x_i^{(1)}(n-1) + x_i^{(0)}(n)), \\ &i = 1, 2, \dots, N, \end{aligned}$$

其中 $x_i^{(1)}(k) = \sum_{j=1}^k x_i^{(0)}(j) (k = 1, 2, \dots, n)$. 取 $x_1^{(1)}$ 的均值(mean) 数列

$$z_1^{(1)}(k) = 0.5x_1^{(1)}(k) + 0.5x_1^{(1)}(k-1) \quad (k = 2, 3, \dots, n),$$

则 $z_1^{(1)} = (z_1^{(1)}(2), z_1^{(1)}(3), \dots, z_1^{(1)}(n))$. 于是可得到 GM(1, N) 的灰微分方程为

$$x_1^{(0)}(k) + ax_1^{(1)}(k) = \sum_{i=2}^N b_i x_i^{(1)}(k), \quad (20.7)$$

其中 $x_1^{(0)}(k)$ 为灰导数, $z_1^{(1)}(k)$ 为背景值, $a, b_i (i = 2, 3, \dots, N)$ 为参数.

如果对于一切时刻 $k = 2, 3, \dots, n$, 引入向量矩阵记号:

$$Y_N = (x_1^{(0)}(2), x_1^{(0)}(3), \dots, x_1^{(0)}(n))^T, u = (a, b_2, b_3, \dots, b_N)^T$$

$$B = \begin{bmatrix} -z_1^{(1)}(2) & x_2^{(1)}(2) & \cdots & x_N^{(1)}(2) \\ -z_1^{(1)}(3) & x_2^{(1)}(3) & \cdots & x_N^{(1)}(3) \\ \vdots & \vdots & & \vdots \\ -z_1^{(1)}(n) & x_2^{(1)}(n) & \cdots & x_N^{(1)}(n) \end{bmatrix},$$

则 GM(1, N) 的灰微分方程为

$$Y_N = B \cdot u$$

其中 Y_N 为已知数据向量, B 为 GM(1, N) 已知数据矩阵, u 为参数向量. 用 \hat{u} 表示 u 的估计值, 令 $\varepsilon = Y_N - B \cdot \hat{u}$ 表示估计值的残差, 根据最小二乘法, 求使

$$J(\hat{u}) = \varepsilon^T \varepsilon = (Y_N - B \cdot \hat{u})^T (Y_N - B \cdot \hat{u})$$

达到最小值的估计值 \hat{u} .

事实上, 如果存在 $(B^T \cdot B)^{-1}$, 则有

$$\hat{u} = (\hat{a}, \hat{b}_2, \hat{b}_3, \dots, \hat{b}_N)^T = (B^T B)^{-1} B^T Y_N. \quad (20.8)$$

如果 $(B^T \cdot B)$ 为奇异矩阵 (例如当 $n-1 < N$ 时), 即 $(B^T \cdot B)^{-1}$ 不存在, 则此时 \hat{u} 不能用 (20.8) 式确定. 但注意到 \hat{u} 的元素实际上是各子因素对主因素影响大小的反映, 因此, 我们可以引入加权矩阵 $W = \text{diag}(w_1, w_2, \dots, w_N)$, 使对各因素的未来发展趋势进行调整控制. 对于未来发展减弱趋势的因素赋予较大的权值, 而对于未来增强趋势的因素赋予较小的权值, 使之更好地反映未来的实际

情况. 则类似地方法可以得到相应参数的估计值

$$\hat{\mathbf{u}} = (\hat{a}, \hat{b}_2, \hat{b}_3, \dots, \hat{b}_N)^T = \mathbf{W}^{-1} \mathbf{B}^T (\mathbf{B}^T \mathbf{W}^{-1} \mathbf{B})^{-1} \mathbf{Y}_N.$$

2. GM(1, N) 的白化型

对于模型 GM(1, N) 的灰微分方程 (20.7), 如果将 $x_i^{(1)}(k)$ 的时刻 $k = 1, 2, \dots, N$ 视为连续变量 t , 则数列 $x_i^{(1)}(k)$ 就可以视为时间 t 的函数, 记为 $x_i^{(1)} = x_i^{(1)}(t)$. 则可得到 GM(1, N) 的白化微分方程

$$\frac{dx_1^{(1)}}{dt} + ax_1^{(1)}(t) = \sum_{i=2}^N b_i x_i^{(1)}(t),$$

即为一阶 N 个变量的微分方程.

20.3 灰色预测

灰色预测是指利用 GM 模型对系统行为特征的发展变化规律进行估计预测, 同时也可以对行为特征的异常情况发生的时刻进行估计计算, 以及对在特定区内发生事件的未来时间分布情况做出研究等等. 这些工作实质上是将“随机过程”当作“灰色过程”, “随机变量”当作“灰变量”, 并主要以灰色系统理论中的 GM(1, 1) 模型来进行处理.

灰色预测在工业、农业、商业等经济领域, 以及环境、社会 and 军事等领域中都有广泛的应用. 特别是依据目前已有的数据对未来的发展趋势做出预测分析.

20.3.1 灰色预测的方法

设已知参考数据列为 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$, 做 1 次累加 (AGO) 生成数列

$$\begin{aligned} x^{(1)} &= (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n)) \\ &= (x^{(0)}(1), x^{(0)}(1) + x^{(0)}(2), \dots, x^{(0)}(n-1) + x^{(0)}(n)), \end{aligned}$$

其中 $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(i)$ ($k = 1, 2, \dots, n$). 求均值 (mean) 数列

$$z^{(1)}(k) = 0.5x^{(1)}(k) + 0.5x^{(1)}(k-1) \quad (k = 2, 3, \dots, n),$$

即 $z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(n))$. 于是建立灰微分方程为

$$x^{(0)}(k) + az^{(1)}(k) = b \quad (k = 2, 3, \dots, n).$$

相应的白化微分方程为

$$\frac{dx^{(1)}}{dt} + ax^{(1)}(t) = b, \quad (20.9)$$

记 $u = (a, b)^T$, $Y_1 = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n))^T$, $B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \vdots & \vdots \\ -z^{(1)}(n) & 1 \end{bmatrix}$, 则由

最小二乘法, 求使得 $J(\hat{u}) = (Y_1 - B \cdot \hat{u})^T (Y_1 - B \cdot \hat{u})$ 达到最小值的 $\hat{u} = (\hat{a}, \hat{b})^T = (B^T B)^{-1} B^T Y_1$. 于是求解方程 (20.9) 得

$$x^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} \quad (k = 1, 2, \dots, n-1). \quad (20.10)$$

20.3.2 灰色预测的步骤

1. 数据的检验与处理

首先, 为了保证建模方法的可行性, 需要对已知数据列做必要的检验处理. 设参考数据为 $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$, 计算数列的级比

$$\lambda(k) = \frac{x^{(0)}(k-1)}{x^{(0)}(k)} \quad (k = 2, 3, \dots, n).$$

如果所有的级比 $\lambda(k)$ 都落在可容覆盖 $X = (e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}})$ 内^[1,3], 则数列 $x^{(0)}$ 可以作为模型 GM(1,1) 和进行数据灰色预测. 否则, 需要对数列 $x^{(0)}$ 做必要的变换处理, 使其落入可容覆盖内. 即取适当的常数 c , 作平移变换

$$y^{(0)}(k) = x^{(0)}(k) + c \quad (k = 1, 2, \dots, n),$$

则使数列 $y^{(0)} = (y^{(0)}(1), y^{(0)}(2), \dots, y^{(0)}(n))$ 的级比

$$\lambda_y(k) = \frac{y^{(0)}(k-1)}{y^{(0)}(k)} \in X \quad (k = 2, 3, \dots, n).$$

2. 建立模型 GM(1,1)

按 20.3.1 节中的方法建立模型 GM(1,1), 则可以得到预测值

$$\hat{x}^{(1)}(k+1) = \left(x^{(0)}(1) - \frac{b}{a} \right) e^{-ak} + \frac{b}{a} \quad (k = 1, 2, \dots, n-1),$$

而且 $\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) \quad (k = 1, 2, \dots, n-1)$.

3. 检验预测值

(1) 残差检验: 令残差为 $\varepsilon(k)$, 计算

$$\varepsilon(k) = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{x^{(0)}(k)} \quad (k = 1, 2, \dots, n),$$

如果 $\varepsilon(k) < 0.2$, 则可认为达到一般要求; 如果 $\varepsilon(k) < 0.1$, 则认为达到较高的要求.

(2) 级比偏差值检验: 首先由参考数据 $x^{(0)}(k-1), x^{(0)}(k)$ 计算出级比

$\lambda_0(k)$,再用发展系数 a 求出相应的级比偏差

$$\rho(k) = 1 - \left(\frac{1 - 0.5a}{1 + 0.5a} \right) \lambda_0(k),$$

如果 $\rho(k) < 0.2$,则可认为达到一般要求;如果 $\rho(k) < 0.1$,则认为达到较高的要求.

4. 预测预报

由模型 GM(1,1)所得到的指定时区内的预测值,实际问题的需要,给出相应的预测预报.

20.4 灰色决策

灰色决策主要包括灰局势决策和灰模式关联决策等.

20.4.1 灰局势决策

1. 局势

设 a_i 为(要做决策)一个事件, b_j 为事件 a_i 的第 j 个对策,则称二元组合 $s_{ij} = (a_i, b_j)$ 为一个局势,其中 $i \in I = \{1, 2, \dots, n\}$, $j \in J = \{1, 2, \dots, m\}$.

如果已知发生事件 a_i ,对于 a_i 有一组对策 $\{b_j\} (j \in J)$,按不同的目标 $p \in P = \{1, 2, \dots, l\}$. 在 $\{b_j\}$ 中选择不同的对策来应对事件 a_i ,则会有不同的效果. 一般说来,我们要试图针对目标 p 在 $\{b_j\}$ 中选择效果最佳的对策来对付 a_i . 于是,我们将“事件、对策、目标和效果”称为决策的四要素.

2. 效果样本矩阵

如果对于事件 a_i 和对策 b_j 有局势 $s_{ij} = (a_i, b_j)$, $i \in I = \{1, 2, \dots, n\}$, $j \in J = \{1, 2, \dots, m\}$. 记局势 s_{ij} 在目标 p 下的效果样本为 $u_{ij}^p (p \in P)$,则称矩阵

$$U = \begin{bmatrix} u_{11}^p & u_{12}^p & \cdots & u_{1m}^p \\ u_{21}^p & u_{22}^p & \cdots & u_{2m}^p \\ \vdots & \vdots & & \vdots \\ u_{n1}^p & u_{n2}^p & \cdots & u_{nm}^p \end{bmatrix}$$

为在目标 p 下的效果样本矩阵. 而称 $u_{ij} = (u_{ij}^1, u_{ij}^2, \dots, u_{ij}^l)$ 为样本序列.

3. 效果测度

对于某一目标而言,如果要求效果样本为极大值,则称此目标为正极性目标. 相反的,如果要求效果样本值为极小值,则称此目标为负极性目标. 如果要求效果样本为适中值,则称此目标为中极性目标. 如何来度量样本的效果呢?

对于效果样本 $u_{ij}^p (p \in P)$,我们用 r_{ij}^p 来表示 u_{ij}^p 的效果测度 ($i \in I, j \in J, p \in$

P). 于是, 根据目标 p 的极性, 定义效果测度如下:

$$(1) \text{ 如果 } u_{ij}^p \text{ 为正极性, 则 } r_{ij}^p = \frac{u_{ij}^p}{\max_i \max_j u_{ij}^p};$$

$$(2) \text{ 如果 } u_{ij}^p \text{ 为负极性, 则 } r_{ij}^p = \frac{\min_i \min_j u_{ij}^p}{u_{ij}^p};$$

$$(3) \text{ 如果 } u_{ij}^p \text{ 为中极性, 则 } r_{ij}^p = \frac{\min\{u_0, u_{ij}^p\}}{\max\{u_0, u_{ij}^p\}}, \text{ 其中 } u_0 \text{ 为适中值.}$$

4. 统一测度与满意对策

实际中, 可以将不同极性的效果样本 u_{ij}^p , 通过某种变换成为统一的具有相同极性的效果测度 $r_{ij}^p (p \in P)$, 则称 r_{ij}^p 的全体为极性一致 (统一) 空间. 对于所有 $p = 1, 2, \dots, l$, 则记 r_{ij}^l 为局势 s_{ij} 的统一效果测度, 或统一测度, 即

$$r_{ij}^l = \frac{1}{l} \sum_{p=1}^l r_{ij}^p \quad (i \in I, j \in J).$$

设 r_{ij}^l 为局势 $s_{ij} = (a_i, b_j)$ 的统一测度, 如果存在 $j^* \in J$ 使得 $r_{ij^*}^l = \max_j r_{ij}^l (i \in I)$, 则称 $s_{ij^*} = (a_i, b_{j^*})$ 为事件 a_i 的满意局势, 称 b_{j^*} 为 a_i 的满意对策.

5. 灰局势决策的计算步骤

按照灰局势决策的一般方法, 可以归纳计算步骤如下:

- (1) 确定决策的四要素: 事件、对策、局势和目标;
- (2) 确定局势效果样本;
- (3) 确定目标的极性, 并通过效果测度变换使其统一极性;
- (4) 计算统一测度;
- (5) 求解满意局势.

20.4.2 灰模式决策

灰模式决策主要包括灰模式的评估与辨识, 目的是确定优化 (或满意) 模式, 即优化方案.

1. 模式指标

设决策问题共有 m 个模式 (方案) $x_i (i = 1, 2, \dots, m)$, 每个模式 x_i 都有 n 个数据指标 $x_i(k) (k = 1, 2, \dots, n)$, 记为

$$x_i = (x_i(1), x_i(2), \dots, x_i(n)) \quad (i = 1, 2, \dots, m),$$

则称之为第 i 个模式指标数列.

对于所有 $k = 1, 2, \dots, n$, 可有

$$x(k) = (x_1(k), x_2(k), \dots, x_m(k)) \quad (k = 1, 2, \dots, n),$$

则称 $x(k)$ 为第 k 个指标模式列, 简称为 k 指标.

实际中的每一个指标都有一个确定的极性,即极大值极性,极小值极性和适中值极性.为了方便,不妨设所有指标均为极大值极性,否则,总可以通过适当的变换转化为极大值极性.对于极大值极性的指标模式列 $x(k)$ ($k=1,2,\cdots,n$),如果

$$x_0(k) = \max_i x_i(k) \quad (k=1,2,\cdots,n),$$

则称

$$x_0 = \{x_0(k)\} = (x_0(1), x_0(2), \cdots, x_0(n))$$

为标准模式.

2. 模式优化度与优化模式

设 $x_i = (x_i(1), x_i(2), \cdots, x_i(n))$ 为第 i ($i=1,2,\cdots,m$) 个模式, $x_0 = \{x_0(k)\}$ 为标准模式,则称 x_i 对于 x_0 的灰关联度

$$r_i = r(x_0, x_i) = \frac{1}{n} \sum_{k=1}^n r(x_0(k), x_i(k))$$

为模式优化度(满意度),记为 $g_0(i)$,即 $g_0(i) = r(x_0, x_i)$ ($i \in I$).

如果存在 $i^* \in I$ 使得

$$g_0(i^*) = r(x_0, x_{i^*}) = \max_{i \in I} r(x_0, x_i),$$

则称 x_{i^*} 为优化模式.

3. 灰模式决策的计算步骤

对于要研究的灰模式决策问题,可以按下列步骤求解优化模式:

- (1) 给出模式指标数列 x_i 及指标极性;
- (2) 构造指标模式列 $x(k)$,并在需要时通过极性变换转化为极大值极性;
- (3) 构造标准模式 x_0 ;
- (4) 计算模式的优化度 $g_0(i)$;
- (5) 求出优化模式 x_{i^*} .

20.5 SARS 疫情对某些经济指标影响问题

20.5.1 问题的提出

2003 年的 SARS 疫情对中国部分行业的经济发展产生了一定的影响,特别是对部分疫情较严重的省市的相关行业所造成的影响是明显的,经济影响主要分为直接经济影响和间接影响.直接经济影响涉及商品零售业、旅游业、综合服务等行业.很多方面难以进行定量地评估,现仅就 SARS 疫情较重的某市商品零售业、旅游业和综合服务业的影响进行定量的评估分析.

究竟 SARS 疫情对商品零售业、旅游业和综合服务业的影响有多大,已知某市从 1997 年 1 月到 2003 年 12 月的商品零售额、接待旅游人数和综合服务收入的统计数据如表 20-1、表 20-2 和表 20-3^[7]。

表 20-1 商品的零售额(单位:亿元)

| 年代 | 1 月 | 2 月 | 3 月 | 4 月 | 5 月 | 6 月 | 7 月 | 8 月 | 9 月 | 10 月 | 11 月 | 12 月 |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1997 | 83.0 | 79.8 | 78.1 | 85.1 | 86.6 | 88.2 | 90.3 | 86.7 | 93.3 | 92.5 | 90.9 | 96.9 |
| 1998 | 101.7 | 85.1 | 87.8 | 91.6 | 93.4 | 94.5 | 97.4 | 99.5 | 104.2 | 102.3 | 101.0 | 123.5 |
| 1999 | 92.2 | 114.0 | 93.3 | 101.0 | 103.5 | 105.2 | 109.5 | 109.2 | 109.6 | 111.2 | 121.7 | 131.3 |
| 2000 | 105.0 | 125.7 | 106.6 | 116.0 | 117.6 | 118.0 | 121.7 | 118.7 | 120.2 | 127.8 | 121.8 | 121.9 |
| 2001 | 139.3 | 129.5 | 122.5 | 124.5 | 135.7 | 130.8 | 138.7 | 133.7 | 136.8 | 138.9 | 129.6 | 133.7 |
| 2002 | 137.5 | 135.3 | 133.0 | 133.4 | 142.8 | 141.6 | 142.9 | 147.3 | 159.6 | 162.1 | 153.5 | 155.9 |
| 2003 | 163.2 | 159.7 | 158.4 | 145.2 | 124 | 144.1 | 157.0 | 162.6 | 171.8 | 180.7 | 173.5 | 176.5 |

表 20-2 接待海外旅游人数(单位:万人)

| 年代 | 1 月 | 2 月 | 3 月 | 4 月 | 5 月 | 6 月 | 7 月 | 8 月 | 9 月 | 10 月 | 11 月 | 12 月 |
|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1997 | 9.4 | 11.3 | 16.8 | 19.8 | 20.3 | 18.8 | 20.9 | 24.9 | 24.7 | 24.3 | 19.4 | 18.6 |
| 1998 | 9.6 | 11.7 | 15.8 | 19.9 | 19.5 | 17.8 | 17.8 | 23.3 | 21.4 | 24.5 | 20.1 | 15.9 |
| 1999 | 10.1 | 12.9 | 17.7 | 21.0 | 21.0 | 20.4 | 21.9 | 25.8 | 29.3 | 29.8 | 23.6 | 16.5 |
| 2000 | 11.4 | 26.0 | 19.6 | 25.9 | 27.6 | 24.3 | 23.0 | 27.8 | 27.3 | 28.5 | 32.8 | 18.5 |
| 2001 | 11.5 | 26.4 | 20.4 | 26.1 | 28.9 | 28.0 | 25.2 | 30.8 | 28.7 | 28.1 | 22.2 | 20.7 |
| 2002 | 13.7 | 29.7 | 23.1 | 28.9 | 29.0 | 27.4 | 26.0 | 32.2 | 31.4 | 32.6 | 29.2 | 22.9 |
| 2003 | 15.4 | 17.1 | 23.5 | 11.6 | 1.78 | 2.61 | 8.8 | 16.2 | 20.1 | 24.9 | 26.5 | 21.8 |

表 20-3 综合服务业累计数额(单位:亿元)

| 年代 | 2 月 | 3 月 | 4 月 | 5 月 | 6 月 | 7 月 | 8 月 | 9 月 | 10 月 | 11 月 | 12 月 |
|------|-----|-----|-----|-----|-----|-------|-------|-------|-------|-------|-------|
| 1997 | 96 | 144 | 194 | 276 | 383 | 466 | 554 | 652 | 747 | 832 | 972 |
| 1998 | 111 | 169 | 235 | 400 | 459 | 565 | 695 | 805 | 881 | 1 011 | 1 139 |
| 1999 | 151 | 238 | 335 | 425 | 541 | 641 | 739 | 866 | 975 | 1 087 | 1 238 |
| 2000 | 164 | 263 | 376 | 531 | 600 | 711 | 913 | 1 038 | 1 173 | 1 296 | 1 497 |
| 2001 | 182 | 318 | 445 | 576 | 708 | 856 | 1 000 | 1 145 | 1 292 | 1 435 | 1 667 |
| 2002 | 216 | 361 | 504 | 642 | 818 | 979 | 1 142 | 1 305 | 1 479 | 1 644 | 1 920 |
| 2003 | 241 | 404 | 584 | 741 | 923 | 1 114 | 1 298 | 1 492 | 1 684 | 1 885 | 2 218 |

试根据这些历史数据建立预测评估模型,评估 2003 年 SARS 疫情给该市的商品零售业、旅游业和综合服务业所造成的影响。

20.5.2 模型的分析与假设

根据所掌握的历史统计数据可以看出,在正常情况下,全年的平均值较好地反映了相关指标的变化规律,这样可以把预测评估分成两部分:

(1) 利用灰色理论建立 GM(1,1) 模型,由 1997—2002 年的平均值预测 2003 年平均值;

(2) 通过历史数据计算每个月的指标值与全年总值的关系,从而可预测出正常情况下 2003 年每个月的指标值,再与实际值比较可以估算出 SARS 疫情实际造成的影响。

给出下面两条假设:

(1) 假设该市的统计数据都是可靠准确的;

(2) 假设该市在 SARS 疫情流行期间和结束之后,数据的变化只与 SARS 疫情的影响有关,不考虑其他随机因素的影响。

20.5.3 建立灰色预测模型 GM(1,1)

由已知数据,对于 1997—2002 年某项指标记为矩阵 $A = (a_{ij})_{6 \times 12}$,计算每年的年平均值,记为

$$x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(6)), \quad (20.11)$$

并要求级比 $\sigma(i) = x^{(0)}(i-1)/x^{(0)}(i) \in (0.7515, 1.3307) (i=2,3,\dots,6)$. 对 $x^{(0)}$ 作一次累加,则

$$x^{(1)}(1) = x^{(0)}(1), x^{(1)}(i) = \sum_{k=1}^i x^{(0)}(k) \quad (i=2,3,\dots,6),$$

记

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(6)). \quad (20.12)$$

取 $x^{(1)}$ 的加权均值,则 $z^{(1)}(k) = \alpha x^{(1)}(k) + (1-\alpha)x^{(1)}(k-1) (k=2,3,\dots,6)$, α 为确定参数,记

$$z^{(1)} = (z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(6)), \quad (20.13)$$

于是 GM(1,1) 的白化微分方程模型为

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b, \quad (20.14)$$

其中 a 是发展灰度, b 是内生控制灰度。

由于 $x^{(1)}(k) - x^{(1)}(k-1) = x^{(0)}(k)$, 取 $x^{(0)}(k)$ 为灰导数, $z^{(1)}(k)$ 为背景值,则将方程 (20.14) 对相应的灰微分方程为

$$x^{(0)}(k) + az^{(1)}(k) = b(k = 2, 3, \dots, 6)$$

或

$$x^{(0)}(k) = -az^{(1)}(k) + b(k = 2, 3, \dots, 6),$$

即矩阵形式为

$$Y^{(0)} = B \cdot (a, b)^T,$$

$$\text{其中 } Y^{(0)} = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(6))^T, B = \begin{pmatrix} -z^{(1)}(2) & -z^{(1)}(3) & \dots & -z^{(1)}(6) \\ 1 & 1 & \dots & 1 \end{pmatrix}^T.$$

用最小二乘法求得参数的估计值为

$$(\hat{a}, \hat{b})^T = (B^T \cdot B)^{-1} \cdot B^T \cdot Y^{(0)}. \quad (20.15)$$

于是方程(20.14)有响应(特解)

$$\hat{x}^{(1)}(t+1) = \left(x^{(0)}(1) - \frac{b}{a}\right) \cdot e^{-at} + \frac{b}{a},$$

则

$$\hat{x}^{(0)}(k+1) = \hat{x}^{(1)}(k+1) - \hat{x}^{(1)}(k) = \left(x^{(0)}(1) - \frac{b}{a}\right) \cdot (e^{-ak} - e^{-a(k-1)}). \quad (20.16)$$

由(20.16)式可以得到2003年的平均值为 \bar{x} , 则预测2003年的总值为 $X = 12 \cdot \bar{x}$. 根据历史数据, 可以统计计算出2003年第 i 个月的指标值占全年总值的比例为 u_i , 即

$$u_i = \sum_{j=1}^6 a_{ij} / \sum_{i=1}^{12} \sum_{j=1}^6 a_{ij} \quad (i = 1, 2, \dots, 12), \quad (20.17)$$

则 $u = (u_1, u_2, \dots, u_{12})$, 于是可得2003年每一个月的指标值为 $Y = X \cdot u$.

20.5.4 模型的求解

(1) 商品零售额

由数据表20-1, 用(20.11)和(20.12)式计算可得年平均、一次累加值分别为

$$\begin{aligned} x^{(0)} &= (87.6167, 98.5000, 108.4750, 118.4167, \\ &\quad 132.8083, 145.4083), \\ x^{(1)} &= (87.6167, 186.1167, 294.5917, 413.0083, \\ &\quad 545.8167, 691.2250). \end{aligned}$$

显然 $x^{(0)}$ 的所有级比都在可容区域内. 经检验, 在这里取参数 $\alpha = 0.4$ 比较

合适,由(20.13)式则有

$$z^{(1)} = (127.016\ 7, 229.506\ 7, 341.958\ 3, 466.131\ 7, 603.980\ 0).$$

由最小二乘法用(20.15)式求得 $a = -0.099\ 3, b = 85.598\ 5$. 由(20.16)式可得2003年的月平均值为 $\bar{x} = 162.882\ 6$ 亿元;年总值为 $X = 12\bar{x} = 1\ 954.6$ 亿元. 由(20.17)式得每月的比例为

$$u = (0.079\ 4, 0.080\ 7, 0.074\ 9, 0.078\ 6, 0.081\ 9, 0.081\ 8, \\ 0.084\ 5, 0.083\ 8, 0.087\ 2, 0.088\ 6, 0.086\ 6, 0.092\ 0),$$

故2003年1—12月的预测值为

$$Y = u \cdot X = (155.2, 157.8, 146.4, 153.6, 160.1, 159.9, \\ 165.2, 163.8, 170.5, 173.2, 169.3, 179.9) \quad (\text{亿元})$$

将预测值与实际统计值进行比较如表20-4所示.

表 20-4 2003 年商品的零售额(单位:亿元)

| 月 份 | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 7月 | 8月 | 9月 | 10月 | 11月 | 12月 |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 预测值 | 155.2 | 157.8 | 146.4 | 153.6 | 160.1 | 159.9 | 165.2 | 163.8 | 170.5 | 173.2 | 169.3 | 179.9 |
| 实际值 | 163.2 | 159.7 | 158.4 | 145.2 | 124.0 | 144.1 | 157.0 | 162.6 | 171.8 | 180.7 | 173.5 | 176.5 |

(2) 接待海外旅游人数

由数据表20-2,用(20.11),(20.12)式计算年平均值 $x^{(0)}$ 和一次累加值 $x^{(1)}$. 取参数 $\alpha = 0.5$, 由(20.13)式可得加权平均值 $z^{(1)}$. 由(20.15),(20.16),(20.17)式可求得 $a = -0.093\ 8, b = 16.267\ 0, \bar{x} = 30.264\ 9, X = 12 \cdot \bar{x} = 363.178\ 8$, 以及

$$u = (0.040\ 7, 0.073\ 2, 0.070\ 3, 0.087\ 8, 0.090\ 7, 0.084\ 8, \\ 0.083\ 6, 0.102\ 2, 0.101\ 0, 0.104\ 1, 0.091\ 4, 0.070\ 1)$$

于是可得到2003年的接待海外旅游人数的预测值,并与实际值比较如表20-5所示.

表 20-5 2003 年接待海外旅游人数(单位:万人)

| 月份 | 1月 | 2月 | 3月 | 4月 | 5月 | 6月 | 7月 | 8月 | 9月 | 10月 | 11月 | 12月 |
|-----|------|------|------|------|------|------|------|------|------|------|------|------|
| 预测值 | 14.8 | 26.6 | 25.5 | 31.9 | 33.0 | 30.8 | 30.4 | 37.1 | 36.7 | 37.8 | 33.2 | 25.5 |
| 实际值 | 15.4 | 17.1 | 23.5 | 11.6 | 1.78 | 2.61 | 8.8 | 16.2 | 20.1 | 24.9 | 26.5 | 21.8 |

(3) 综合服务业累计数额

首先将表20-3的数据进行必要的处理,然后用(20.11),(20.12)式计算年平均值 $x^{(0)}$ 和一次累加值 $x^{(1)}$. 取参数 $\alpha = 0.4$, 由(20.13)式可得加权平均值 $z^{(1)}$. 用(20.15),(20.16),(20.17)式可求得 $a = -0.136\ 3, b = 84.155\ 7, \bar{x} =$

203.788 5, $X = 11 \cdot \bar{x} = 2\,241.7$, 2003 年月比例数和每个月的综合服务业的收入额分别为

$$u = (0.109\,1, 0.067\,9, 0.070\,7, 0.090\,2, 0.078\,1, 0.084\,1, 0.097\,8, 0.091\,1, 0.087\,3, 0.089\,9, 0.133\,8)$$

$$Y = X \cdot u = (244, 153, 158, 202, 175, 188, 219, 204, 196, 201, 300).$$

于是可得到 2003 年的综合服务业累计数额的预测值, 并与实际值比较如表 20-6 所示.

表 20-6 2003 年综合服务业累计数额(单位: 亿元)

| 月份 | 2 月 | 3 月 | 4 月 | 5 月 | 6 月 | 7 月 | 8 月 | 9 月 | 10 月 | 11 月 | 12 月 |
|-----|-----|-----|-----|-----|-----|-------|-------|-------|-------|-------|-------|
| 预测值 | 244 | 397 | 555 | 757 | 933 | 1 121 | 1 340 | 1 445 | 1 740 | 1 942 | 2 242 |
| 实际值 | 241 | 404 | 584 | 741 | 923 | 1 113 | 1 298 | 1 492 | 1 684 | 1 885 | 2 218 |

20.5.5 模型的结果分析

根据该市的统计报告显示, 2003 年 4、5、6 三个月的实际商品零售额分别为 145.2、124、144.1 亿元. 在这之前, 根据统计部门的估计 4、5、6 三月份 SARS 疫情对该市的商品零售业的影响最为严重, 这三个月估计大约损失 62 亿元左右. 从我们的模型预测结果来计算, 4、5、6 三个月的损失为 60.3 亿元, 这个数据基本与专家的估算值相符, 8 月份基本恢复正常, 这也说明了模型的正确性和可靠性.

对于旅游业来说是受影响最严重的行业之一, 最严重的 4、5、6、7 四个月就损失 100 多万人, 按最新统计数据^[1], 平均每人消费 1002 美元计算, 大约损失 10 亿美元. 全年大约损失 160 万人, 约合 16 亿美元, 到年底基本恢复正常.

对于综合服务业中的部分行业影响较大, 如航空交通运输、宾馆餐饮等, 但有些行业影响不大, 如电信、通讯等, 总平均来看, 影响还不算太大, 5、6、7、8 四个月大约损失 70 亿元.

从预测结果可以看出, 虽然下半年没有发生疫情, 但人们一直担心 SARS 会卷土重来, 所以, 对这些行业还是有一定的影响的, 即 SARS 影响的延续性的作用.

该模型虽是就某经济指标的发展规律进行评估预测而建立的, 但类似地也适用于其他方面的一些数据规律的评估预测问题, 即该模型具有很广泛的应用性.

20.6 参考案例与参考文献

1. 参考案例

(1) 出版物差错数的预测问题——文献[1]: 358—363

- (2) 100 米跑的成绩预测问题——文献[2]:235—236
- (3) 合理购房问题——文献[3]:190—196
- (4) 深海电缆的参数设计问题——文献[3]:196—204
- (5) 瓷砖制造工艺的优化问题——文献[3]:228—235
- (6) 输油管道防腐蚀问题——文献[3]:236—241
- (7) 事故伤亡率问题——文献[5]:58—63
- (8) 胜利渠的灌溉问题——文献[4]:472—487

2. 参考文献

- [1] 邓聚龙. 灰色系统理论教程. 武汉:华中理工大学出版社,1990
- [2] 沈继红等. 数学建模. 哈尔滨:哈尔滨工程大学出版社,1998
- [3] 邓聚龙. 灰预测与灰决策. 武汉:华中科技大学出版社,2002
- [4] 邓聚龙. 灰色控制系统. 第二版. 武汉:华中理工大学出版社,1993
- [5] 赵云胜等. 灰色系统理论在地学中的应用研究. 武汉:华中理工大学出版社,1997
- [6] 夏军. 灰色系统水文学. 武汉:华中理工大学出版社,2000
- [7] 北京市统计信息网:<http://www.bjstats.gov.cn/>

本书比较全面、系统和简练地介绍了常用的 20 类数学建模方法，每一类都自成体系；此外，每种方法都附有新颖、生动的应用实例，全书涉及面广，表述简练，实用性强，充分体现了“广、浅、精、用”的现代应用科学教育的特点。

本书的主要内容包括量纲分析、集合分析、微分方程、差分方程、赋值与拟合、层次分析、概率分布、数理统计、回归分析、线性规划、整数规划、非线性规划、动态规划、排队论、对策论、随机性决策分析、多目标决策分析、图论、模糊数学和灰色系统分析等方法，最后附有历年中国和美国大学生数学建模竞赛题以及 MATLAB 优化和统计两个工具箱的使用简介。本书的应用实例包括儿童保险、黄河小浪底调水调沙、足球门的危险区域、沼气的生成、雨水北调的水指标分配、招聘公务员、选拔队员与组队、校园网的设计与调节收费、玫瑰有约、彩票中的数学、股份制公司的综合投资、中介服务机构的信誉评估等问题。

本书除可作为高等院校校本、专科生“数学建模”课程教材外，还特别适用于数学建模竞赛的培训教材，以及供从事应用研究的工程技术人员参考。

Methods of Mathematical Modeling

9 787040 163285 >



ISBN 7-04-016328-4

定价 31.10 元

查阅其他教学用书，敬请浏览

www.boson.com.cn

